
MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
on: Thu Dec 4 16:03:31 1997; MasPar time 8.83 Seconds
614.595 Million cell updates/sec
lar output not generated.
Title: >US-08-790-043A-1
Description: (1-256) from US08790043A.pap
Perfect Score: 1753
Sequence: 1 MLNLENTYVINGIANKRSI.....ISSGVGTGENIHVDSGFHAIK 256
Scoring table: PAM 150
Gap 11
Searched: 59021 seqs, 21210388 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: swiss-prot34
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11
Statistics: Mean 49.533; Variance 114.207; scale 0.434
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
on: Thu Dec 4 16:03:31 1997; MasPar time 8.83 Seconds
614.595 Million cell updates/sec
lar output not generated.
Title: >US-08-790-043A-1
Description: (1-256) from US08790043A.pap
Perfect Score: 1753
Sequence: 1 MLNLENTYVINGIANKRSI.....ISSGVGTGENIHVDSGFHAIK 256
Scoring table: PAM 150
Gap 11
Searched: 59021 seqs, 21210388 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: swiss-prot34
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11
Statistics: Mean 49.533; Variance 114.207; scale 0.434
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES

Hit No.	Score	Query Match	Length	ID	Description	Pred. No.
1	865	49.3	264	3	FABI_ANASP PUTATIVE ENOYL-[ACYL-	3.73e-125
2	748	42.7	261	3	FABI_ECOLI ENOYL-[ACYL-CARRIER-P	2.09e-104
3	732	41.8	261	3	FABI_SALTY ENOYL-[ACYL-CARRIER-P	1.39e-101
4	718	41.0	261	3	FABI_HAEN ENOYL-[ACYL-CARRIER-P	4.06e-99
5	365	20.8	385	3	FABI_BRANA ENOYL-[ACYL-CARRIER-P	1.29e-38
6	357	20.4	269	5	INHA_MYCTU ENOYL-[ACYL-CARRIER-P	2.64e-37
7	342	19.5	269	5	INHA_MYCTU ENOYL-[ACYL-CARRIER-P	7.37e-35
8	229	13.1	249	1	BA71_EUBSP 7-ALPHA-HYDROXYSTEROI	4.16e-17
9	225	12.8	249	1	BA72_EUBSP 7-ALPHA-HYDROXYSTEROI	1.65e-16
10	219	12.5	244	1	AP27_MOUSE ADIPOCYTE P27 PROTEIN	1.29e-15
11	220	12.5	260	9	TRN2_HYONI TROPINONE REDUCTASE-I	9.16e-16
12	214	12.2	260	9	TRN2_DATST TROPINONE REDUCTASE-I	7.07e-15
13	197	11.2	245	6	NODG_RHIME MODULATION PROTEIN G	2.12e-12
14	183	10.4	254	11	YJGU_ECOLI HYPOTHETICAL OXIDORED	2.07e-10
15	179	10.2	255	4	HDHA_ECOLI 7-ALPHA-HYDROXYSTEROI	7.54e-10
16	176	10.0	262	11	YXAU_BACSU GLUCONATE 5-DEHYDROGE	1.97e-09
17	171	9.8	256	4	GNO_GLOOX HYPOTHETICAL OXIDORED	9.64e-09
18	171	9.8	320	3	FABG_CUPLA 3-OXOACYL-[ACYL-CARRI	1.81e-08
19	169	9.6	261	11	YJGI_ECOLI HYPOTHETICAL OXIDORED	9.64e-09
20	168	9.6	268	9	TRNH_DATST TROPINONE REDUCTASE H	2.48e-08
21	167	9.5	246	6	NODG_AZOBTR MODULATION PROTEIN G	3.40e-08
22	166	9.5	278	1	ARDH_PICST D-ARABINITOL 2-DEHYDR	4.64e-08

23	164	9.4	244	3	FABG_ECOLI 3-OXOACYL-[ACYL-CARRI	8.66e-08
24	163	9.3	250	6	LINX_PSEPA 2,5-DICHLORO-2,5-CYCL	1.18e-07
25	163	9.3	256	10	VER1_ASPPA VERSICOLORIN REDUCTAS	1.18e-07
26	162	9.2	278	4	FIXR_BRAJA FIXR PROTEIN	1.61e-07
27	159	9.1	295	9	SPF19_YEAST SPORULATION PROTEIN S	4.07e-07
28	156	8.9	114	11	YHXC_BACSU HYPOTHETICAL OXIDORED	1.02e-06
29	156	8.9	140	11	YHXC_BACSU HYPOTHETICAL OXIDORED	1.02e-06
30	156	8.9	250	6	LINC_PSEPA 2,5-DICHLORO-2,5-CYCL	1.02e-06
31	155	8.8	273	9	TRN1_DATST TROPINONE REDUCTASE-I	1.38e-06
32	155	8.8	281	1	ARDH_CANAL D-ARABINITOL 2-DEHYDR	1.38e-06
33	153	8.7	244	3	FABG_VIBHA 3-OXOACYL-[ACYL-CARRI	2.54e-06
34	152	8.7	254	5	KDUD_BACSU 2-DEOXY-D-GLUCONATE 3	3.43e-06
35	153	8.7	267	4	HDHA_CLOSO 7-ALPHA-HYDROXYSTEROI	2.54e-06
36	153	8.7	282	1	ARDH_CANTR D-ARABINITOL 2-DEHYDR	2.54e-06
37	150	8.6	257	11	YXJF_BACSU HYPOTHETICAL OXIDORED	6.26e-06
38	148	8.4	242	3	FABG_HAEN 3-OXOACYL-[ACYL-CARRI	1.14e-05
39	147	8.4	248	7	PHAB_ACISP ACETOACETYL-COA REDUC	1.53e-05
40	147	8.4	255	11	YHFD_BACSU HYPOTHETICAL OXIDORED	2.78e-05
41	145	8.3	247	10	Y07E_MYCTU PUTATIVE OXIDOREDUCTA	1.20e-04
42	140	8.0	246	7	PHBB_ALCEU ACETOACETYL-COA REDUC	1.20e-04
43	140	8.0	289	8	PTRI_LEITA PTERIDINE 1 REDUCTASE	1.61e-04
44	139	7.9	249	3	DHK2_STRVN GRANATICIN POLYKETIDE	1.61e-04
45	138	7.9	261	3	DHG2_BACME GLUCOSE 1-DEHYDROGENA	2.15e-04

ALIGNMENTS

RESULT 1	FABI_ANASP	STANDARD;	PRT;	264 AA.
AC Q05069;				
DT 01-FEB-1994 (REL. 28, CREATED)				
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)				
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)				
DE PUTATIVE ENOYL-[ACYL-CARRIER-PROTEIN] REDUCTASE (NADH) (EC 1.3.1.9)				
DE (NADH-DEPENDENT ENOYL-ACP REDUCTASE).				
OS ANABAENA SP. (STRAIN PCC 7120).				
OC PROKARYOTA; GRACILICUTES; OXYPHOTOBACTERIA;				
OC CYANOBACTERIA (BLUE-GREEN ALGAE); NOSTOCALES.				
RN [1]				
RP SEQUENCE FROM N.A.				
RA WEI T.-F., RAMASUBRAMANIAN T.S., PU F., GOLDEN J.W.;				
RL J. BACTERIOL. 175:4025-4035(1993).				
CC -!- CATALYTIC ACTIVITY: ACYL-[ACYL-CARRIER PROTEIN] + NAD(+) =				
CC TRANS-2,3-DEHYDROACYL-[ACYL-CARRIER PROTEIN] + NADH.				
CC -!- PATHWAY: SECOND REDUCTIVE STEP IN FATTY ACID BIOSYNTHESIS.				
CC -!- SIMILARITY: SOME SIMILARITY TO THE SHORT-CHAIN				
CC DEHYDROGENASES/REDUCTASES FAMILY (SDR).				
DR EMBL: L10036; G142010;				
KW HYPOTHETICAL PROTEIN; OXIDOREDUCTASE; NAD; FATTY ACID BIOSYNTHESIS.				
FT NP_BIND 10 36 NAD (BY SIMILARITY).				
SQ SEQUENCE 264 AA; 28083 MW; A78ADCB CRC32;				

Query Match	49.3%	Score 865;	DB 3;	Length 264;
Best Local Similarity	51.6%	Pred. No. 3.73e-125;		
Matches	133;	Conservative	58;	Mismatches 62;
				Indels 5;
				Gaps 5;
Db	7	mlnltgknaivtgiannrsiaiaqqlhaaganlitylpd-ergkfekkvselvepln	65	
QY	1	MLNLENTYVINGIANKRSIAGFVAKVLDQLGAKLVFTYRKRSRKELEKLLSQLNP-E	59	
Db	66	pslflpncvndeqiqstfdti-rdkwgrldilhlclafanrdldtqdsqtsragfata	124	
QY	60	AHLQ-QIDVQSDSEVINGFQIGKDV-GNIDGVHSHSTAFANMDELGRFSETSEGFLLA	117	
Db	125	ldistfsivlgkaakplmteggssitlslvgvrvaypnynvmgvakagleasvrylase	184	
QY	118	QDISYSYITVAHEAKKLMPGGSSIVATYLGGEFAVONYNVMGVAKEANVKVYALD	177	
Db	185	lgsqnrivnaissagpirtlassavvgldmhvveqvaprrrtvtqlvengtfaasdl	244	
QY	178	LGPDIRVNAISSAGPIRTLSAKVGGFNTILKEIEERAPLKRNVYDQVEVGKTAAYLLSDL	237	

4 LENKIIIVIMGLANKRRSIAFGVAKVLDQLGAKLVFTYRKERSRKEKLEQLNQPEAHLY 63

102 NIKVNAISAGPIRTLSAKGVGGFNTILKEIEERAPLKRNVQVEVGKTAAYLLSDLSS

Fri Dec 5 12:31:32 1997

US-08-790-043A-1.rsp

```

Db 240 sgevhvdgff 250
QY 242 TGENHVDSGF 252

RESULT 4
ID FABI_HAEIN STANDARD; PRT; 261 AA.
AC P44432;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DE 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ENOYL-[ACYL-CARRIER-PROTEIN] REDUCTASE (NADH) (EC 1.3.1.9) (NADH-
DE DEPENDENT ENOYL-ACP REDUCTASE).
GN FABI OR ENVM OR H11734.
OS HAEMOPHILUS INFLUENZAE.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC PASTURELLACEAE.
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=RD / KW20;
RX MEDLINE; 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA UTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.;
RA SCIENCE 269:496-512(1995).
RL ENOYL-[ACYL-CARRIER PROTEIN] + NAD(+) =
CC -1- CATALYTIC ACTIVITY: ACYL-[ACYL-CARRIER PROTEIN] + NADH.
CC TRANS-2,3-DEHYDROACYL-[ACYL-CARRIER PROTEIN] + NADH.
CC -1- PATHWAY: SECOND REDUCTIVE STEP IN FATTY ACID BIOSYNTHESIS.
CC -1- THE ANTIBIOTIC DIAZABORINE INTERFERES WITH THE ACTIVITY BY BINDING
CC TO THE PROTEIN.
CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE-ASSOCIATED (BY SIMILARITY).
CC -1- SIMILARITY: SOME SIMILARITY TO THE SHORT-CHAIN
CC DEHYDROGENASES/REDUCTASES FAMILY (SDR).
DR EMBL; U32846; G926824; ALT_INIT.
KW OXIDOREDUCTASE; NAD; FATTY ACID BIOSYNTHESIS; ANTIBIOTIC RESISTANCE;
KW INNER MEMBRANE.
FT INIT_MET 0 0 BY SIMILARITY.
FT NP_BIND 9 35 NAD (BY SIMILARITY).
FT SEQUENCE 261 AA; 27988 MW; EFPA381B CRC32;

Query Match 41.0%; Score 718; DB 3; Length 261;
Best Local Similarity 43.2%; Pred. No. 4.06e-99;
Matches 108; Conservative 70; Mismatches 69; Indels 3; Gaps 3;

Db 3 ltkgrilvtglasnrslaygiaksmkqgaelfaftylnkdklqprveefakegf-sdivl- 60
QY 4 LENKTVINGIAKRSIAFGVAKVLDQGLAKLVFTYRKERSKLEKLEQLNQPEAHLY 63
Db 61 plvatdeslqncfaelskrwdkfdgfhafapgdldgdynaatreyyriahdisa 120
QY 64 QIDVQSDVEEVINGFEIGDKVDGNDGVHYSIAFANMEDLRGRF-SETSGREGFLAQDISS 122
Db 121 ysfvamaqaarpylnpnaalltlylgaerainpynvncmlakasleaatrvmadlqkeg 180
QY 123 YSLTIYAHEAKKLMPGGSGSVIATYTGGEFAVONTNMGVAKASLEANVKYALDLDPDN 182
Db 181 irvnaisagpirtlaasgiknfmkmlstfekaalrtvttiedvgnsaafclsdlasgit 240
QY 183 IRVNAISAGPIRLSAKGVGGFTNLKEIERAPLKRNVQDQVEVGKTAAYLLSDLSGVT 242
Db 241 geivhdagf 250
QY 243 GENHVDSGF 252

RESULT 5
ID FABI_BRANA STANDARD; PRT; 385 AA.
AC P44432;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DE 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ENOYL-[ACYL-CARRIER-PROTEIN] REDUCTASE (NADH) (EC 1.3.1.9) (NADH-
DE DEPENDENT ENOYL-ACP REDUCTASE).
GN FABI OR ENVM OR H11734.
OS HAEMOPHILUS INFLUENZAE.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
OC PASTURELLACEAE.
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=RD / KW20;
RX MEDLINE; 95350630.
RA FLEISCHMANN R.D., ADAMS M.D., WHITE O., CLAYTON R.A., KIRKNESS E.F.,
RA KERLAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICK J.M.,
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAYNE J.D.,
RA SCOTT J.D., SHIRLEY R., LIU L.-I., GLODEK A., KELLEY J.M.,
RA WEIDMAN J.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.D.,
RA UTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDEK D.M., BRANDON R.C.,
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA GNEHM C.L., MCDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.,
RA VENTER J.C.;
RA SCIENCE 269:496-512(1995).
RL ENOYL-[ACYL-CARRIER PROTEIN] + NAD(+) =
CC -1- CATALYTIC ACTIVITY: ACYL-[ACYL-CARRIER PROTEIN] + NADH.
CC TRANS-2,3-DEHYDROACYL-[ACYL-CARRIER PROTEIN] + NADH.
CC -1- PATHWAY: SECOND REDUCTIVE STEP IN FATTY ACID BIOSYNTHESIS.
CC -1- THE ANTIBIOTIC DIAZABORINE INTERFERES WITH THE ACTIVITY BY BINDING
CC TO THE PROTEIN.
CC -1- SUBCELLULAR LOCATION: INNER MEMBRANE-ASSOCIATED (BY SIMILARITY).
CC -1- SIMILARITY: SOME SIMILARITY TO THE SHORT-CHAIN
CC DEHYDROGENASES/REDUCTASES FAMILY (SDR).
DR EMBL; U32846; G926824; ALT_INIT.
KW OXIDOREDUCTASE; NAD; FATTY ACID BIOSYNTHESIS; ANTIBIOTIC RESISTANCE;
KW INNER MEMBRANE.
FT INIT_MET 0 0 BY SIMILARITY.
FT NP_BIND 9 35 NAD (BY SIMILARITY).
FT SEQUENCE 261 AA; 27988 MW; EFPA381B CRC32;

Query Match 41.0%; Score 718; DB 3; Length 261;
Best Local Similarity 43.2%; Pred. No. 4.06e-99;
Matches 108; Conservative 70; Mismatches 69; Indels 3; Gaps 3;

Db 3 ltkgrilvtglasnrslaygiaksmkqgaelfaftylnkdklqprveefakegf-sdivl- 60
QY 4 LENKTVINGIAKRSIAFGVAKVLDQGLAKLVFTYRKERSKLEKLEQLNQPEAHLY 63
Db 61 plvatdeslqncfaelskrwdkfdgfhafapgdldgdynaatreyyriahdisa 120
QY 64 QIDVQSDVEEVINGFEIGDKVDGNDGVHYSIAFANMEDLRGRF-SETSGREGFLAQDISS 122
Db 121 ysfvamaqaarpylnpnaalltlylgaerainpynvncmlakasleaatrvmadlqkeg 180
QY 123 YSLTIYAHEAKKLMPGGSGSVIATYTGGEFAVONTNMGVAKASLEANVKYALDLDPDN 182
Db 181 irvnaisagpirtlaasgiknfmkmlstfekaalrtvttiedvgnsaafclsdlasgit 240
QY 183 IRVNAISAGPIRLSAKGVGGFTNLKEIERAPLKRNVQDQVEVGKTAAYLLSDLSGVT 242
Db 241 geivhdagf 250
QY 243 GENHVDSGF 252

RESULT 6
ID INHA_MYCSM STANDARD; PRT; 269 AA.
AC P42829;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DE 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ENOYL-[ACYL-CARRIER-PROTEIN] REDUCTASE (NADH) (EC 1.3.1.9) (NADH-
DE DEPENDENT ENOYL-ACP REDUCTASE).
GN INHA.
OS MYCOBACTERIUM SMEGMATIS.
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; MYCOBACTERIACEAE.
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=MC SQUARED 155;

```

[illegible]

Fri Dec 5 12:31:32 1997

```

DR PIR: B37762; B37762.
DR HSSP; P19992; LHDC.
DR PROSITE; PS00061; ADH_SHORT.
KW OXIDOREDUCTASE; NAD; BILE ACID CATABOLISM; MULTIGENE FAMILY.
FT NP_BIND 7 13 NAD (POTENTIAL).
FT NP_BIND 32 36 NAD (POTENTIAL).
FT ACT_SITE 157 157 BY SIMILARITY.
FT CONFLICT 159 162 TSKA -> YOGG (IN REF. 2).
SQ SEQUENCE 249 AA; 26658 MW; 9E49D579 CRC32;

Query Match 13.1%; Score 229; DB 1; Length 249;
Best Local Similarity 23.7%; Pred. No. 4.16e-17;
Matches 61; Conservative

Db 1 mklvqdktiltg-gt-rqigfaaakliengakvsiifgetqeevdtalaql-kelypee 57
QY 2 LNL-ENKTVIMGIANKRSIAFGVAKVLDQGLAKL-VFTYRKSRKELEKLLQLNQPE 59

Db 58 evlqfapdltsrdavmaavgtvqkygrldvminnagit-mnsfsvrseefkn-imdi 115
60 AHL-YQIDVQSDVEVINGEQIGKDVGNIDGVYHSIAFANMEDLGRFSETSREGFLAQ 118
116 nvngvngawsqcmk-dakqviintastvgisgigypstkgavlgthlgre 174
QY 119 DISS-YSLTIVAHEAKKLMPEGGSIVATTYLGGEFAVQNNVMGVAKASLEANKVYLALD 177
175 iirknirrvqavgvvtdmtkqlpp-e-iledyktlpmkmlkpeelanvylflasdl 232
QY 178 LGPDNIRVNAISAGPIRTLSAKGVGGFNILKEIERAPLKRNVQVEVGKTAAYLLSDL 237
233 asgittattisvsgayrtp 249
QY 238 SSGVTGENIHVDSGFHA 254

RESULT 9 STANDARD; PRT; 249 AA.
AC P19337;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE 7-ALPHA-HYDROXYSTEROID DEHYDROGENASE (EC 1.1.1.159) (BILE ACID
DE 7-DEHYDROXYLASE) (7-ALPHA-HSDH) (BILE ACID-INDUCIBLE PROTEIN).
GN BAI2.
OS EUBACTERIUM SP. (STRAIN VFI 12708).
OC PROKARYOTA; FIRMICUTES; IRREGULAR ASPOROGENOUS RODS;
OC PROPIONIBACTERIACEAE.
[1]
SEQUENCE FROM N.A.
MEDLINE: 91072253.
RA MALONEE D.H., WHITE W.B., HYLEMON P.B.;
RL J. BACTERIOL. 172:7011-7019(1990).
[2]
SEQUENCE FROM N.A.
MEDLINE: 90330548.
RA GOPAL-SRIVASTAVA R., MALLONEE D.H., WHITE W.B., HYLEMON P.B.;
RL J. BACTERIOL. 172:4420-4426(1990).
[3]
SEQUENCE FROM N.A.
MEDLINE: 90264339.
RA MALONEE D.H., WHITE W.B., HYLEMON P.B.;
RL J. BACTERIOL. 172:3400-3408(1990).
[4]
SEQUENCE FROM N.A.
MEDLINE: 89008068.
RA WHITE W.B., FRANKLUND C.V., COLEMAN J.P., HYLEMON P.B.;
RL J. BACTERIOL. 170:4555-4561(1988).
CC -1- FUNCTION: 7-ALPHA-DEHYDROXYLATION OF CHOLIC ACID, YIELDING
CC DEOXYCHOLIC ACID AND LITHOCHOLIC ACID, RESPECTIVELY. HIGHEST
CC AFFINITY WITH TAUROCHENOXYCHOLIC ACID.
CC CATALYTIC ACTIVITY: 3-ALPHA, 7-ALPHA, 12-ALPHA-TRIHYDROXY-5-BETA-
CC CHOLANATE + NAD(+) = 3-ALPHA, 12-ALPHA-DIHYDROXY-7-EXO-5-BETA-
CC CHOLANATE + NADH.

CC -1- PATHWAY: BILE ACID METABOLISM.
CC -1- INDUCTION: PRESENCE OF C(24) BILE ACIDS CONTAINING A
CC 7-ALPHA-HYDROXY GROUP.
CC -1- SIMILARITY: THERE ARE THREE GENES FOR BAI2 PROTEINS: BAI1 IS
CC IDENTICAL TO BAI3 AND THERE IS 8% IDENTITY WITH BAI2.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
DR EMBL; M36292; G148528; -
DR EMBL; M26223; G290676; -
DR EMBL; U57489; G1381567; -
DR PIR; A31841; A31841.
DR PIR; E37844; E37844.
DR HSSP; P19992; LHDC.
DR PROSITE; PS00061; ADH_SHORT.
KW OXIDOREDUCTASE; NAD; BILE ACID CATABOLISM; MULTIGENE FAMILY.
FT NP_BIND 7 13 NAD (POTENTIAL).
FT NP_BIND 32 36 NAD (POTENTIAL).
FT ACT_SITE 157 157 BY SIMILARITY.
SQ SEQUENCE 249 AA; 26538 MW; 453C0028 CRC32;

Query Match 12.8%; Score 225; DB 1; Length 249;
Best Local Similarity 23.7%; Pred. No. 1.65e-16;
Matches 61; Conservative

Db 1 mnlvqdktiltg-gt-rqigfaaakliengakvsiifgetqeevdtalaql-kelypee 57
QY 2 LNL-ENKTVIMGIANKRSIAFGVAKVLDQGLAKL-VFTYRKSRKELEKLLQLNQPE 59

Db 58 evlqfapdltsrdavmaavgtvqkygrldvminnagit-mnsfsvrseefk-himdi 115
60 AHL-YQIDVQSDVEVINGEQIGKDVGNIDGVYHSIAFANMEDLGRFSETSREGFLAQ 118
116 nvngvngawsqcmk-dakqviintastvgisgigypstkgavlgthlgre 174
QY 119 DISS-YSLTIVAHEAKKLMPEGGSIVATTYLGGEFAVQNNVMGVAKASLEANKVYLALD 177
175 iirknirrvqavgvvtdmtagnpp-e-imegykkaipmkmlkpeelanvylflasdl 232
QY 178 LGPDNIRVNAISAGPIRTLSAKGVGGFNILKEIERAPLKRNVQVEVGKTAAYLLSDL 237
233 asgittattisvsgayrtp 249
QY 238 SSGVTGENIHVDSGFHA 254

RESULT 10 STANDARD; PRT; 244 AA.
AC P08074;
DT 01-AUG-1988 (REL. 08, CREATED)
DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ADIPOCYTE P27 PROTEIN (AP27).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
[1]
SEQUENCE FROM N.A.
RX STRAIN-CH3;
RX MEDLINE: 88273310.
RA NAVRE M., RINGOLD G.M.;
RL J. CELL BIOL. 107:279-286(1988).
CC -1- INDUCTION: THIS PROTEIN IS INDUCED BY GLUCOCORTICOID.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC FAMILY (SDR).
DR EMBL; D26123; G699608; -
DR EMBL; X07411; G50004; -
DR PIR; A28053; A28053.
DR PIR; S03382; S03382.
DR PROSITE; PS00061; ADH_SHORT.
KW OXIDOREDUCTASE.
FT NP_BIND 11 35 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 149 149 BY SIMILARITY.
SQ SEQUENCE 244 AA; 25958 MW; D8C37C9B CRC32;

```

<hr/>					
RESULT	12				
ID	TRN2_DATST	STANDARD;	PRT;	260 AA.	
AC	P50163;				
DT	01-OCT-1996	(REL. 34, CREATED)			
DT	01-OCT-1996	(REL. 34, LAST SEQUENCE UPDATE)			
DT	01-OCT-1996	(REL. 34, LAST ANNOTATION UPDATE)			
DE	TROPINONE REDUCTASE-II	(EC 1.1.1.236) (TR-II).			
OS	DATURA STRAMONIUM	(JIMSONWEED) (COMMON THORNAPPLE).			
OC	EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;				
OC	SOLANALES; SOLANACEAE.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=ROOT;				
RA	MEDLINE; 94022421.				
RX	NAKAJIMA K., HASHIMOTO T., YAMADA Y.;				
RL	PROC. NATL. ACAD. SCI. U.S.A. 90:9591-9595(1993).				
CC	-!- FUNCTION: CATALYSES THE STEREOSPECIFIC REDUCTION OF TROPINONE TO PSEUDOTROPINE.				
CC	-!- CATALYTIC ACTIVITY: PSEUDOTROPINE + NADP(+) = TROPINONE + NADPH.				
CC	-!- PATHWAY: BIOSYNTHETIC PATHWAY OF TROPANE ALKALOIDS.				
CC	-!- SUBUNIT: HOMOTETRAMER (PROBABLE).				
CC	-!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR).				
DR	EMBL; L20474; G424162; -.				
KW	OXIDOREDUCTASE; NADP.				
FT	NP_BIND 13 37				
FT	ACT_SITE 159 159				
FT	BY SIMILARITY.				
SC	SEQUENCE 260 AA; 28311 MW; 29555C68 CRC32;				
<hr/>					
Query Match					
Best Local Similarity 12.2%; Score 214; DB 9; Length 260;					
Matches 63; Conservative 82; Mismatches 97; Indels 15; Gaps					
<hr/>					
Db	6	nlegctalvtg-gs-rigivgiiveelasigas-vvycsrndkelndcltqwrskgfkea	62		
Qy	3	NLENKTIVVINGIAKRKSIAFGVAKVLQLGAKLVFFY-RKRSRKE-LEKLLEQLNQPEA	60		
Db	63	svcdlsrsrgerelmtvanhf-h-qklnlvmnagiviyeakd-ytvved-yslimsin	118		
Qy	61	HLQYDVQSD-EVINGFBOIGKDVGNDGVVHSTAFAANNEDLRGFSETSGREFLLAQD	119		
Db	119	feayhlsvlahpfk-asergnvvfissvgalavpyeavygatkgamdqltrclaefw	177		
Qy	120	I-SYSLSITVAHEAKKLMPGGSIATTVTLGEFAVQNYVMGVAKASLEANVKYLALDL	178		
Db	178	akdnirvnvggvviatslvemtqidpekenlkldrcalmrgepkeelaamvaficf	237		
Qy	179	GPDNIIRVNIAISAGPIPT-LSAKGVGFN-TI-LKEIEERAPLKRVVDQVEVGKTAAYLLS	235		
Db	238	paasyvtgqiyyvdggsl	254		
Qy	236	DLSSGVTGENIHWDGSF	252		
<hr/>					
RESULT	13				
ID	NODC_RHIME	STANDARD;	PRT;	245 AA.	
AC	P06234;				
DT	01-JAN-1988	(REL. 06, CREATED)			
DT	01-JAN-1988	(REL. 06, LAST SEQUENCE UPDATE)			
DE	01-OCT-1996	(REL. 34, LAST ANNOTATION UPDATE)			
DE	MODULATION PROTEIN G	(HOST-SPECIFICITY OF MODULATION PROTEIN C).			
NO	RDG OR HNSC.				
OS	RHIZORIUM MELIOITI.				
OC	PLASMID SYN.				
OC	PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;				
OC	RHIZOBIACEAE.				
RN	[1]				
RN	SEQUENCE FROM N.A.				
RC	STRAIN=2011;				
RC	MEDLINE; 87016382				
TX					

RL NUCLEIC ACIDS RES. 14:7453-7472(1986).
 [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-1021.
 RA FISHER R.F., SWANSON J.A., MULLIGAN J.T., LONG S.R.;
 RL GENETICS 117:191-201(1987).
 CC -1- FUNCTION: PROPOSED TO MODIFY MOD FACTOR FATTY ACYL CHAIN.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC FAMILY (SDR).
 DR ENBL; X04379; G46308; -.
 DR ENBL; Y00604; G46304; -.
 DR PIR; C24706; C24706.
 DR PIR; S07675; S07675.
 DR HSP; P19992; 1HDC.
 DR PROSITE; PS00061; ADH_SHORT.
 KW MODULATION; OXIDOREDUCTASE; PLASMID; NAD.
 FT NP_BIND 11 35
 FT ACT_SITE 152 152 BY SIMILARITY.
 SQ SEQUENCE 245 AA; 26088 MW; 92F585C0 CRC32;
 Query Match 11.2%; Score 197; DB 6; Length 245;
 Best Local Similarity 24.1%; Pred. No. 2.12e-12;
 Matches 62; Conservative 73; Mismatches 108; Indels 14; Gaps 13;
 Db 1 mfe1grkalvtgag--aiggaiaarvlhaqga-iv-glhtgqiek-letlatelgdr-v 54
 QY 1 MLNLENTYVIMGTANKRSIAFGVAKVLDQLGAKLVFTYRKERSRKELEKLLQNLQPEA 60
 Db 55 klpanlanrdevkalgraseadlegvdilvnnagik-dglflhmadpawd-ivlevnl 112
 QY 61 HLYQIDVQSDDEVINGFEQIGKDVGNIDGVYHSIAFANMEDLRGFRSETSGEFLLAQDI 120
 Db 113 tamfirtre1tg-qm1rringriinvtvagaipnpgntnycaskagm1gfsk1aqla 171
 QY 121 SS-YSLTIVAHEAKLMPEGGSIVATYVLGGFAVQNVNMGVAKASLEANKVYALDLIG 179
 Db 172 trnitvncvapgfi-e-sam-tklnhkqekimvaipihmrgtgvavasavaylasdha 228
 QY 180 PDNIRVNAISAGPIRTLSAKGVGGFNTILKE-IEERAPLKRNVQDVEVGKTAAYLLSDLS 238
 Db 229 avytgqtlhvggmami 245
 QY 239 SGVTGENIHVDSGFHAI 255
 RESULT 14
 ID YJGU_ECOLI STANDARD; PRT; 254 AA.
 P39345;
 01-FEB-1995 (REL. 31, CREATED)
 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL OXIDOREDUCTASE IN PEPA-GNTV INTERGENIC REGION
 DE (EC 1.-.-.-) (F234).
 GN YJGU.
 OS ESCHERICHIA COLI.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 OC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE; 95334362.
 RA BURLAND V.D., PLUNKETT G. III, SOFIA H.J., DANIELS D.L.,
 RA BLATTNER F.R.;
 RL NUCLEIC ACIDS RES. 23:2105-2119(1995).
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC FAMILY (SDR).
 DR ENBL; U14003; G537108; -.
 DR ECOGENE; EG12540; YJGU.
 DR PROSITE; PS00061; ADH_SHORT.
 KW HYPOTHETICAL PROTEIN; OXIDOREDUCTASE.
 FT NP_BIND 13 37 NAD OR NADP (BY SIMILARITY).
 FT ACT_SITE 158 158 BY SIMILARITY.
 SQ SEQUENCE 254 AA; 27563 MW; 908F3A66 CRC32;

Query Match 10.4%; Score 183; DB 11; Length 254;
 Best Local Similarity 21.8%; Pred. No. 2.07e-10;
 Matches 55; Conservative 76; Mismatches 114; Indels 7; Gaps 7;
 Db 4 lfs1agknillitg-sa-ggigflilatgkygagilinditaeraelaveklhgeigav 61
 QY 1 MLNLENTYVIMGTANKRSIAFGVAKVLDQLGAKLVFTYRKERSRKELEKLLQNLQPE 59
 Db 62 aapfnv-thkhe-idaavehiekdipdvlnnag1qrrhpfte-fpeqewndvianvq 118
 QY 60 AHLQIDVQSDDEVINGFEQIGKDVGNIDGVYHSIAFANMEDLRGFRSETSGEFLLAQD 119
 Db 119 tavf-lvsqavtrhmverkakvinicmsqselgrdtitpvaaskgavkmltrgmvela 177
 QY 120 ISSYSLTIVAHEAKLMPEGGSIVATYVLGGFAVQNVNMGVAKASLEANKVYALDLG 179
 Db 178 rhnqvngiagpyfktemkalvedeaftawlckrtptarwgdqgeligaavflsskasd 237
 QY 180 PDNIRVNAISAGPIRTLSAKGVGGFNTILKEIEERAPLKRNVQDVEVGKTAAYLLSDLS 239
 Db 238 fvnghllfvddg 249
 QY 240 GVTGENIHVDSG 251
 RESULT 15
 ID HDHA_ECOLI STANDARD; PRT; 255 AA.
 AC P25529;
 DT 01-NAY-1992 (REL. 22, CREATED)
 DT 01-NAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
 DE 7-ALPHA-HYDROXYSTEROID DEHYDROGENASE (EC 1.1.1.159) (7-ALPHA-HSDH).
 GN HDHA OR HSDH.
 OS ESCHERICHIA COLI.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 OC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-HB101;
 RX MEDLINE; 91177803.
 RA YOSHIMOTO T., HIGASHI H., KANATANI A., LIN X.S., NAGAI H., OYAMA H.,
 RA KURAZONO K., TSURU D.;
 RL J. BACTERIOL. 173:2173-2179(1991).
 RN [2]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
 RX MEDLINE; 96264882.
 RA TANAKA N., NONAKA T., TANABE T., YOSHIMOTO T., TSURU D., MITSUI Y.;
 CC BIOCHEMISTRY 35:7715-7730(1996).
 CC -1- FUNCTION: 7-ALPHA-DEHYDROXYLATION OF CHOLIC ACID, YIELDING
 CC DEOXYCHOLIC ACID AND LITHOCHOLIC ACID, RESPECTIVELY. HIGHEST
 CC AFFINITY WITH TAUROCHENODEOXYCHOLIC ACID.
 CC -1- CATALYTIC ACTIVITY: 3-ALPHA,7-ALPHA,12-ALPHA-TRIHYDROXY-5-BETA-
 CC CHOLANATE + NAD(+) = 3-ALPHA,12-ALPHA-DIHYDROXY-7-OXO-5-BETA-
 CC CHOLANATE + NADH.
 CC -1- SUBUNIT: HOMOTETRAMER.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC FAMILY (SDR).
 DR ENBL; D10497; G216571; -.
 DR PIR; A38527; A38527.
 DR PIR; J0951; J0951.
 DR HSP; P19992; 1HDC.
 DR ECOGENE; EG10425; HDHA.
 DR PROSITE; PS00061; ADH_SHORT.
 KW OXIDOREDUCTASE; NAD; BILE ACID CATABOLISM.
 FT NP_BIND 18 24 NAD.
 FT ACT_SITE 159 159
 SQ SEQUENCE 255 AA; 26778 MW; 443F6382 CRC32;
 Query Match 10.2%; Score 179; DB 4; Length 255;
 Best Local Similarity 25.7%; Pred. No. 7.54e-10;
 Matches 65; Conservative 74; Mismatches 101; Indels 13; Gaps 11;

Db 7 lridgkcailtg-aga-gigkeiaitfatagasvvvsdinadaanhvvdaiqqlg-gqaf 63
 QY 2 LNLENTYVINGIANKRSIAFGVAKVLDQLCAKLVFTYRKERSRKELEKLEQLNQPEAH 61
 Db 64 acrcditsegealsadfa-isk-lgkvdiivnnaggggpkpfmdpmadfr-ayelnv- 119
 QY 62 LYQIDVQSDREV--INGFEQIGKDVGNIDGVYHSIAFANMEDLGRFSETSRGFLIAQD 119
 Db 120 fsffhlsqlyvapemek--ngggviltitsmaeenkninmtsyasskaaaashlvrmnafdl 177
 QY 120 ISSYSLT-IVAHEAKKLMPGGGIVATYLGGEFAVQYNNVMGVAKASLEANKVYLALDL 178
 Db 178 geknirvngiapgailtdalksvit-peieqkmlqhtpirrlgqpqdi anaalfcspaa 236
 QY 179 GPDNIRVNAISAGPIRTLSAKGVGGFNTILKEIERAPLKRNVQVEVGKTAAYLLSDLS 238
 Db 237 swvsqqltvsgg 249
 QY 239 SGVTGENIHVDSG 251

Job completed: Thu Dec 4 16:04:55 1997
 Job time : 84 secs.

Priebe
790643
seq.lds
Interf.

M P E R L H
(TM)

Release 2.1D John F. Collins, BioComputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
on: Thu Dec 4 16:00:39 1997; MasPar time 8.20 Seconds
384.011 Million cell updates/sec
Mular output not generated.

Title: >US-08-790-043A-1
Description: (1-256) from US08790043A.pep
Perfect Score: 1753
Sequence: 1 MLNLENTYVINGIANKRSI.....LSSGVGTGNIHVDSGFHAIK 256

Scoring table: PAM 150
Gap 11

Searched: 101610 seqs, 12294212 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq28
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21

Statistics: Mean 33.197; Variance 180.878; scale 0.184

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	365	20.8	385	4	R23793 Stearoyl-ACP-desaturase	1.25e-17
2	355	20.3	269	12	R66293 M. smegmatis InhA.	6.84e-17
3	355	20.3	269	12	R66289 M. smegmatis InhA.	6.84e-17
4	340	19.4	269	12	R66290 M. tuberculosis InhA.	8.65e-16
5	340	19.4	269	12	R66901 M. tuberculosis InhA.	8.65e-16
6	317	18.1	231	12	R66292 Mycobacterium bovis I	4.15e-14
7	317	18.1	231	12	R63900 M. bovis InhA.	4.15e-14
8	174	9.9	255	4	R22993 7-alpha-hydroxy stero	5.28e-04
9	171	9.8	256	20	W02111 Glucuronate:NADP+-5-oxi	8.41e-04
10	150	8.6	315	15	R89323 Rape leaf beta-ketoac	2.09e-02
11	150	8.6	315	15	R89322 Rape seed beta-ketoac	2.09e-02
12	140	8.0	246	6	R21974 Acetoacetyl CoA reduc	9.35e-02
13	140	8.0	246	6	R32192 Sequence encoded by t	9.35e-02
14	140	8.0	246	1	P94157 Acetyl-CoA reductase.	9.35e-02
15	140	8.0	329	13	R71324 N-acetyl mannosamine	1.69e-01
16	136	7.8	269	3	R14145 NAD affinity glucosine	3.05e-01
17	132	7.5	261	5	R27756 Glucose dehydrogenase	4.73e-01
18	131	7.5	261	5	R27757 Acetoacetyl CoA reduc	4.73e-01
19	129	7.4	246	12	R74761 Glucose dehydrogenase	4.09e-01
20	130	7.4	261	1	P80063	

21	128	7.3	261	1	P80590	Sequence of glucose d	5.47e-01
22	127	7.2	241	2	R10679	Acetoacetyl CoA reduc	6.33e-01
23	118	6.7	247	12	R66291	Mycobacterium bovis p	2.31e-00
24	118	6.7	247	12	R63899	M. bovis PS5 ORF1 pro	2.31e-00
25	118	6.7	261	2	R40404	Glucose dehydrogenase	2.31e-00
26	118	6.7	261	2	R24018	Thermostable glucose	2.31e-00
27	108	6.2	240	13	R61477	Clavulanic acid dehyd	9.47e-00
28	108	6.2	240	13	R77866	S. clavuligerus ORF9	9.47e-00
29	108	6.2	273	1	P94155	Acetyl-CoA reductase.	9.47e-00
30	107	6.1	261	1	R03846	Modified glucose dehy	1.09e+01
31	105	6.0	418	4	P40134	Sequence of human alp	1.44e+01
32	103	5.9	394	12	R67362	Alpha-1-antitrypsin m	1.89e+01
33	103	5.9	394	2	R03754	Entire sequence of co	1.89e+01
34	103	5.9	394	12	R67363	Alpha-1-antitrypsin m	1.89e+01
35	103	5.9	394	12	R67360	Human alpha-1-antitry	1.89e+01
36	103	5.9	394	12	R67361	Alpha-1-antitrypsin m	1.89e+01
37	103	5.9	399	2	R04033	GAPDH promoter fragme	1.89e+01
38	104	5.9	418	3	P50877	Sequence encoded by h	1.65e+01
39	103	5.9	418	4	P50021	Sequence of alpha-1-a	1.89e+01
40	103	5.9	418	4	R29331	Alpha-1 antitrypsin c	1.89e+01
41	101	5.8	256	20	W10220	Streptococcal CAMP fa	2.48e+01
42	101	5.8	394	3	P61713	[Phe358] alaphal-antit	2.48e+01
43	101	5.8	394	3	P61710	[Ile358] alaphal-antit	2.48e+01
44	101	5.8	394	3	P61712	[Leu358] alaphal-antit	2.48e+01
45	101	5.8	394	3	P61708	[Gly358] alaphal-antit	2.48e+01

ALIGNMENTS

RESULT 1
ID R23793 standard; Protein; 385 AA.
AC R23793;
DT 04-NOV-1992 (first entry)
DE Stearoyl-ACP-desaturase (from clone pDE57).
KW Delta-9-desaturase; stearyl-ACP; oleoyl-ACP; lipid; biosynthesis.
OS Brassica napus var. Rafel.
FH Key Location/Qualifiers
FT Peptide 1..73
FT /label= sig_peptide
FT Protein 74..385
FT /label= mat_protein
PN NL9002130-A.
PD 16-APR-1992.
PF 28-SEP-1990; 002130.
PR 28-SEP-1990; NL-002130.
PA (TEWE-) STICHTING TECH WETENSCHAPPEN.
DR WPI: 92-157514/19.
DR N-PSDB: Q24482.
PT Cruciferous stearyl-ACP-desaturase coding sequences - for
PT modifying lipid biosynthesis in plants, esp. oilseed rape
PS Disclosure; Fig 2; 31pp; Dutch
CC Stearoyl-ACP-desaturase catalyses the conversion of stearyl-ACP
CC into oleoyl-ACP. Vectors contg. the sequence encoding this enzyme
CC have been used to transform petunia and oilseed rape plants using
CC Agrobacterium tumefaciens, resulting in altered lipid biosynthesis
CC and modification of lipid composition of the plant.
SQ Sequence 385 AA;

Query Match 20.8%; Score 365; DB 4; Length 385;
Best Local Similarity 38.1%; Pred. No. 1.25e-17;
Matches 67; Conservative 54; Mismatches 49; Indels 6; Gaps 5;
Db 199 dfgsidilvhalan-ngepskplleatkrkylaaissayfsyllsh-flpimpgga 255
QY 83 DVGNIDGVYHSTAFANMEDLRGRSEISREGFLAQDISSYSL-TIVAAEAKLMEGGSS 141
Db 256 sisltviaseirliopygggmssakaalesdtrvlafaeagrknirvntisagpglgsraak 315
QY 142 IVATYILGGEFAVQNTYV-MGVAKASLEANYKYLALDLGP-DNIRVNAISAGPRTISAK 199
Db 316 afgldtmeysynnapikqkltadevnaaafvslasaitgatiyvdnglsm 371
QY 200 GVGGFNTILKEIEERAPLKRNVQVEVGKTAIYLLDSSGVGTGENIHDVSGFHAH 255

RESULT 2

IID R6293 standard; Protein; 269 AA.
AC R66293;
AC 03-JUL-1995 (first entry)
DE M. smegmatis Inha.
DT Isoniazid; isonicotinic acid hydrazide; INH; inhA gene; vaccine.
KW Mycobacterium smegmatis.
W09426765-A.
W09426765-A.
24-NOV-1994.
PD 13-MAY-1994; U05398.
PD 13-MAY-1993; NZ-247620.
PR 14-MAY-1993; US-062409.
PR 31-MAR-1994; US-221742.
PR (AGRE-), AGRESEARCH.
PR (BANE//) BANERJEE A.
PR (COLL//) COLLINS D.
PR (JACO//) JACOBS W R.
PR (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
PR (WILS//) WILSON T M.
PR Banerjee A, Collins D, De LISLE GW, Jacobs WR, Wilson TM;
PR WPI; 95-006691/01.
PR N-PSDB; Q75517.
PR Polynucleotide(s) determining mycobacterial resistance to
T isoniazid - useful in diagnosis, treatment and prevention of
T mycobacterial infection, e.g. tuberculosis.
T Disclosure; Fig. 4A-4B; 104pp; English.
C The gene from Mycobacterium smegmatis encoding Inha (R66293),
C the target of action for isoniazid, was identified, isolated,
C cloned and sequenced (Q75517). Mutant inhA genes have been used
C for recombinant vaccine development.
C Sequence 269 AA;
C Q

Query Match	20.3%	Score 355;	DB 12;	Length 269;
Best Local Similarity	35.0%;	Pred. No. 6,84e-17;		
Matches	92;	Conservative 60;	Mismatches 92;	Indels 19; Gaps 14;
Db	5	legkrilvtgiitdsiafhakvacaagaelvlt-gfdrl-klvkriadi-rpkp-apll	61	
QY	4	LENKTVVINGIANKRSIATGAVKVLQDLGAKLVFTYRKRSRKEKLEQLQEQPEAHLY	63	
Db	62	elvdqneehstladritaeigegnkidvrvhaigfmpqsgmginpffdadvedvskgh	121	
QY	64	QIDVQSDSEEVINGFEQIGKDWG--N-IDGVYHSIAFANMEDL--RGRFSETSGEFLLAQD	119	
Db	122	isaysyasiakavlpimnpvggigvmdf-dprrampaynmvkvaksalesvnrfvareag	180	
QY	120	ISSYSLTIIVAHEAKKLMPGSGSIVATTYLGSEFAVQNTNMGVAKASLEANNVLIALLDLG	179	
Db	181	kygvrsnlvaapirrtlamsaivvgalgdeagqgmllgeegwdqraplgwnmkdptpvak	240	
QY	180	PDNIRVNAISAGPIRTLSAKG-VGG-F-T-----ILKEI-EERAPLKRNV-DQVEVGK	228	
Db	241	tvcallsdwlpattgtviyadgg	263	
QY	229	TAAYLKSLDSSGVTGENTHVDSG	251	

RESULT	3
ID	R66289 standard; Protein; 269 AA.
AC	R66289;
DT	03-JUL-1995 (first entry)
DE	M. smegmatis InhA.
DD	Isoniazid; isonicotinic acid hydrazide; INH; inhA gene;
KW	mycobacteria; vaccine.
KW	Mycobacterium smegmatis.
OS	Mycobacterium smegmatis.
PN	W09426312-A.
PD	24-NOV-1994.
PF	12-MAY-1994; U05344.
PR	13-MAY-1993; NZ-247620.
PR	14-MAY-1993; U0-062409.
PR	31-MAR-1994; U5-221742.

PA	(BANE/) BANERJEE A.	
PA	(COLL/) COLLINS D M.	
PA	(DLIS/) DE LISLE G W.	
PA	(JACO/) JACOBS W R.	
PA	(WILS/) WILSON T M.	
PA	(AGRE-) AGRESEARCH.	
PA	(COLL/) COLLINS D.	
PA	(YESH) UNIV YESHIVA EINSTEIN COLLEGE.	
PI	Banerjee A, Collins DM, De Lisle GW, Jacobs WR;	
PI	Wilson TM, Collins D;	
PI	WPI; 95-006366/01.	
DR	N-PSDB; Q78913.	
DR	Gene target for isonicotinic acid hydrazide - used to develop	
PT	prods for diagnosis, treatment, prevention and studies involving	
PT	mycobacterial infections	
PS	Disclosure: Fig 4; 76pp; English.	
CC	Genes from Mycobacterium smegmatis, Mycobacterium tuberculosis and	
CC	Mycobacterium bovis that encode InhA, the target of action for	
CC	isoniazid, were identified, isolated and cloned. Sequences of	
CC	the 3 genes are given in Q78913-15, and encoded amino acids in	
CC	R6289-91. Mutant genes have been used in recombinant vaccine	
CC	development.	
SQ	Sequence 269 AA;	

Query Match	20.3%;	Score 355;	DB 12;	Length 269;
Best Local Similarity	35.0%;	Pred. No. 6.84e-17;		
Matches	92;	Conservative	60;	Mismatches 92;
			Indels	19;
			Gaps	14;
Db	5	legkrilvtgiitdsia	hiakvagaaelvlt-gfdrl-klvtriadrlpkp-apll	61
QY	4	LENTTVVINGIANKRSIA	FGVAKVLDUGAKLVFTYRKSRKELSKLLQLNQPEAHLY	63
Db	62	elvdvqneehstladrtae	igeknidgvyhaigfmpqsgmginpffdadvedvskgh	121
QY	64	QIDVQSDVEEVINGFOIG	KDVG--N-IDGVYHSTAFANMEDL-RGRFSTSRGFLLAQD	119
Db	122	isaysiaslakavlpim	ppgggiivgmdf-dptcampaynmvtvaksalesvrfrvaeq	180
QY	120	ISSYSLTIVIAHEAKKL	IMPEGGSIVATVYLGGEFAVQNYVMGVAKASLSEANVYIALDLG	179
Db	181	kvgvrsnlvaagpirtlams	aivggalgdaeqgmllsegwdqraplgwnmkdptpvak	240
QY	180	PDNIRVNAISAGPITLSAK	-VGG-F-T-----ILKEI-ERAPLKRNV-DOVEYVK	228
Db	241	tvcallsdwlpattgtvi	adgg	263
QY	229	TAAVLLSDSSGVTGENTH	VDVG	251

RESULT	4
ID	R66290 standard; Protein; 269 AA.
AC	R66290;
DT	03-JUL-1995 (first entry)
DE	M. tuberculosis inhA gene.
KW	Isoniazid; isonicotinic acid hydrazide; INH; inhA gene;
OS	vaccine; mycobacteria; ds.
KS	Mycobacterium tuberculosis.
PN	W09426312-A.
PP	24-NOV-1994.
PPF	12-MAY-1994; U05344.
PPR	13-MAY-1993; N2-247620.
PR	14-MAY-1993; US-082409.
PR	31-MAR-1994; US-221742.
PPA	(BANE/) BANERJEE A.
PPA	(COLL/) COLLINS D M.
PPA	(DLIS/) DE LISLE G W.
PPA	(JACO/) JACOBS W R.
PPA	(WILS/) WILSON T M.
PPA	(AGRE-) AGRESEARCH.
PPA	(COLL/) COLLINS D.
PPA	(YESH) UNIV YESHIVA EINSTEIN COLLEGE.
PPA	Banerjee A, Collins DM, De LISLE GW,
PPA	Wilson TM, Collins D;
PPA	Jacobs WR;

Matches	87;	Conservative	68;	Mismatches	92;	Indels	19;	Gaps	14
Db	5	ldgkrilvsgiltssiafhiarvaeqgaqlvlt-gfdrlrl-igrtdrl-pakapll	61						
		: : : : : : : : : : : : : : : : : : : :							
QY	4	LENKTYVIMGIANKRSIAFGVAKVLDQGLAKLVFTYRKRSRKELEKLEQLNQPEAHLY	63						
Db	62	eldvqneehlasagrvtteagagnklgdvvhsgifmpqtgmginpffdapadvskgih	121						
		: : : : : : : : : : : : : : : : : : : : :							
QY	64	QIDVQSDSEVINGFEQIGKDVG--N-IDGVYHSIAFANMEDL-RGRFSETSRGFLLAQD	119						
Db	122	isaygmfdprallimpvgsgivgmfd-dpsrampaynmtvaksalesnrvfvarcag	180						
		: : : : : : : : : : : : : : : : : : : :							
QY	120	ISYSLSITVAHEAKKLPESGSI VATTYLGGEFAVQNYNMGVAKASLEAFNYKLLALDLG	179						
Db	181	kygvrnlvgagpirtlamsaivgalgeaagaqiglleegwdqrapigwmkdatpvak	240						
		: : : : : : : : : : : : : : : : : : : : : : : :							
QY	180	PDNIRVNAISAGPIRTLSAKG-VGG-F--WT--I--LKEI-EERAPLKRNV-DQVEVGK	228						
Db	241	tvcallsdwlpattgdliiyadggaht	266						
		: : : : : : : : : : : : : : : : : : : :							
QY	229	TAAYLSDLSGGVTGENTHVDSEFHA	254						
RESULT	6								
ID	R66292	standard; Protein; 231 AA.							
AC	R66292;								
DT	03-JUL-1995	(first entry)							
DE	Mycobacterium bovis InHa.								
KW	isoniazid; isonicotinic acid hydrazide; INH; InHa gene; pS5 gene;								
OS	vaccine; mycobacteria.								
KS	Mycobacterium bovis.								
PN	W09426312-A.								
PD	24-NOV-1994.								
PF	12-MAY-1994;	U05344.							
PR	13-MAY-1993;	N2-247620.							
PR	14-MAY-1993;	US-062409.							
PR	31-MAR-1994;	US-221742.							
PA	(BANE/) BANERJEE A.								
PA	(COLL/) COLLINS D M.								
PA	(DLIS/) DE LISLE G W.								
PA	(JACO/) JACOBS W R.								
PA	(WILS/) WILSON T M.								
PA	(AGRE-) AGRESEARCH.								
PA	(COLL/) COLLINS D.								
PA	(YESH/) UNIV YESHIVA EINSTEIN COLLEGE.								
PI	Banerjee A, Collins DM, De LISLE GW, Jacobs WR;								
PI	Wilson TM, Collins D;								
DR	WPI: 95-006366/01.								
DR	N-FSDB: Q78915.								
PT	Gene target for isonicotinic acid hydrazide - used to develop								
PT	prods for diagnosis, treatment, prevention and studies involving								
PT	mycobacterial infections								
PS	Disclosure; Fig. 9A-9D; 76pp; English.								
CC	A cosmid library from M. bovis G4/100 was prepared in shuttle								
CC	vector pYU18 and transformed into Mycobacterium smegmatis ms2155.								
CC	The smallest plasmid obtained which conferred resistance to InHa,								
CC	The target of action for isoniazid, was designated p5. The								
CC	sequence of pS5 was determined (Q78915), revealing 2 large open								
CC	reading frames, ORF2 being the InHa gene encoding the protein given								
CC	in R66292.								
SC	Sequence	231 AA;							
QY	Query Match	18.1%;	Score	317;	DB	12;	Length	231;	
	Best Local Similarity	32.2%;	Pred. No.	4.15e-14;					
	Matches	66;	Conservative	53;	Mismatches	77;	Indels	9;	Gaps
Db	5	ldgkrilvsgiltssiafhiarvaeqgaqlvlt-gfdrlrl-igrtdrl-pakapll	61						
	: : : : : : : : : : : : : : : : : : : :								
QY	4	LENKTYVIMGIANKRSIAFGVAKVLDQGLAKLVFTYRKRSRKELEKLEQLNQPEAHLY	63						
Db	62								

Db	122	isaysasmakallipmnpogsgvmdf-dpsrampaynwmvtaksalesvnrifvareag	180
QY	120	ISSYSLRIVAHEAKKLMPEGGSVATYLLGGFEAFQYNVNVGAKSLEANKVYLLALDUG	179
Db	181	kygrvsnlvaagpirtlamsalgg	205
QY	180	PDNIRVNAISAGPRTLSAKGV-GG	203

RESULT

ID R63900 standard; Protein; 231 AA.
AC R63900;
TAC 03-JUL-1995 (first entry)
DDE M. bovis Inha.
KWW Isoniazid; isonicotinic acid hydrazide; INH; inhA gene; p85; vaccine.
KWW Mycobacterium bovis G4/100.
KWW W09426765-A.
KWW 24-NOV-1994.
KWW 13-MAY-1994; U05398.
KWW 13-MAY-1993; N2-247620.
KWW 14-MAY-1993; US-062409.
KWW 31-MAR-1994; US-221742.
KWW (AGRE-) AGRESEARCH.
KWW (BANE-) BANERJEE A.
KWW (COLL-) COLLINS D.
KWW (JACO-) JACOBS W R.
KWW (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
KWW (WILS-) WILSON T M.
KWW Banerjee A, Collins D, De LISLE GW, Jacobs WR, Wilson TM;
KWW WPI; 95-006691/01.
KWW N-PSDB; Q75519.
KWW Polynucleotide(s) determining mycobacterial resistance to
KWW isoniazid - useful in diagnosis, treatment and prevention of
KWW mycobacterial infection, e.g. tuberculosis.
KWW Disclosure; Page 21; 104pp; English.
KWW A cosmid library of Mycobacterium bovis G4/100 was made in vector
KWW pYU818 and transferred into Mycobacterium smegmatis mc2155. The
KWW smallest plasmid obtained which conferred an isoniazid resistance
KWW phenotype on M. smegmatis was designated p85. Sequencing of p85
KWW revealed 2 open reading frames. ORF2 encoded Inha (R63899),
KWW the target of action for isoniazid. Mutants of the inhA gene
KWW have been used for recombinant vaccine development.
KWW Sequence 231 AA;

Query Match	18.1%;	Score 317;	DB 12;	Length 231;
Best Local Similarity	32.2%;	Pred. NO. 4.15e-14;		
Mismatches	66;	Conservative	53;	Mismatches 77; Indels 9; Gaps 8;

5 ldgkrlvsgitdssiafhiaavaeqgaqlvit-gfdrlrl-igrtdrl-pakapil 61
4 LENKTYVIMZANKRSTAFGAKVLDQGLAKLVFYKRSRKELEKLEQLNQPEAHLY 63
62 eldvneehlaslagrvteaigaunkldgvvhaigfmpgtgmgingffdapvadvskgh 121
64 QIDVQSDSEVINGEQIGKDVGG--N-IDGVYHSIAFANMEDL-RGRFSTSRGFFLLAQD 119
122 isayyaamakallipimppggsivgmddf-dpsrampaynvmtrakaalesvrfvareag 180
120 ISSYSLTIVAHEANKLMPEGGSIVATTYLGGEFAVONVMGVAKASLEANYKYLALDLG 179
181 kygvrsnlvaagpirtiamsamlgg 205
180 PDNIRVNAISAGPVRTLSAKGV--GG 203

RESULTS

ID R22993 standard; Protein; 255 AA.
AC R22993;
DE 19-AUG-1992 (first entry)
KW 7-alpha-hydroxy steroid dehydrogenase.
KW Biochemical reagent.
OS Escherichia coli.

Key	Location/Qualifiers
Misc_difference 34	
FT	/note= "unknown amino acid"
FT	J04058883-A.
PN	25-FEB-1992.
PD	
PP	22-JUN-1990. 165391.
PR	22-JUN-1990; JP-165391.
PA	(NIKA-) SHIN NIPPON KAGAKU.
WPI	92-118055/15.
DR	N-PSDB; Q23902.
DR	N-PSDB; Q23902.
PT	7-Alpha-hydroxy steroid dehydrogenase prepn. - by extracting
PT	enzyme from E. coli obtd. by transformation
PS	Claim 1; Fig 2; 7pp; Japanese.
CC	The gene for 7 alpha-hydroxy steroid dehydrogenase was introduced
CC	into a suitable plasmid and used to transform E. coli. The
CC	transformant can express a large amt. of the enzyme. The
CC	transformant is cultivated in culture medium at pH 7.0 at 37 deg. C
CC	under aerobic conditions, then the cells collected, lysed and
CC	fractionated with ammonium sulphate. The resuspended pellet was
CC	desalted and passed over an ion exchange column to recover purified
CC	enzyme. The enzyme can be used as a biochemical reagent.
CC	Sequence 255 AA;
SQ	

Query Match	9.9%	Score 174;	DB 4;	Length 255;
Best Local Similarity	26.0%	Pred. No. 5.28e-04;		

Matches	66:	Conservative	73:	Mismatches	100:	Indels	15:	Gaps	13:
Db	7	lrdgkcalitg-aga-gigkeiaiaifatkxgasvsvvdsinadaanhvdeiqqlg-gqaf	63						
QY	2	LNLENKTVIMGIANKRSIAFGAKVLDQGLAKLFTYRKERSRKELEKLLLEQLNQPEAH	61						
Db	64	acrcditseqelsaladfa-isk-lgkvdiilvnpvdigkxp-fdmpmadfir-ayelnv	119						
QY	62	LYQIDVQSDDEV--INGFEQICKOVGNIDG-VYHSIAFANMEDJRGRETSRSGFFLLAQ	118						
Db	120	-fsffhlsqlvapenek--ngggvilltitsmaeenkninmtsasyasskaashlvrmafd	176						
QY	119	DISSYSLSL-IVAHSAKLMPGGSGIVATYLGGEFAVQNYVMVGAKASLEANKYKIALD	177						
Db	177	lgexnrvnrgapgaailtdalksvit-peieqmlqhtpirrlgppqdianaalfcspa	235						
QY	178	LGPDIRVNAISAGPIRTLSAKVGGGNTILKEIEERAPLKRNVDOVEVGKTAAYLLSDL	237						
Db	236	aswsvsgqiltvsgg	249						
QY	238	SSGVGTGENIHVDSG	251						

RESULT

W02111 standard; Protein; 256 AA.
W02111;
06-MAR-1997 (first entry)
Glucunate:NADP+-5-oxidoreductase.
Glucunate: NADP+: oxidoreductase; Gluconobacter oxydans;
ascorbic acid; tartaric acid; ds.
S Gluconobacter oxydans.
S EP-726320-AA.
D 14-AUG-1996.
D 07-FEB-1996; 101776.
F 07-FEB-1995: DE-003946.
A (RHEI-) RHEIN BIOTECH GES NEUE BIOTECHNOLOGISCHE.
R Bringer-Meyer S, Hollenberg CP, Klases R, Sahm H;
I WPI: 96-364397/37.
R N-PSDB: T36145.
R N-PSDB: T36145.
T 5-keto:gluconate prodrn. by increasing expression of gluconate:
T NADP+-5-oxide:reductase gene - esp. by increasing copy no. in
T Gluconobacter, used as intermediate for ascorbic and tartaric acids
T Claim 8; Page 10-12; 15pp; German.
S The gluconobacter oxydans gene may be used to transform cells,
S to produce higher levels of gluconate:NADP+-5-oxidoreductase.
S Tartaric acid can now be produced without the difficult
S purification involved in prepn. from tartar.
S Sequence 256 AA.

QY	69	SDEEVFEQIGKDV-GNIDGVYHSAFANNMEDLRGFSETSEGFLLAQDSSYSLSLT	127
Db	190	qa-atkimmkrkriiniasvvgllgnigqaaayaaakagvigfsktaaregasrnnvn	248
QY	128	VAHEAKKLMP-EGSIVATTYLGGEFAVONYNVMGVAKASLEANNVKYLLDLGPDNIRVN	186
Db	249	vvcpgfiasdmtak-lge-d-mekkilgtligrvgqpedvagivelfalspaasyitgq	305
QY	187	ALSAGPIRT-LSAKVGVGFTILKEITERAPLKRNVQDEVGKTAAYL-LSDLSGGVTGE	244
Db	306	aftidggi	313
QY	245	NIHVDSGF	252

RESULT	11	
ID	R89322	standard; Protein; 315 AA.
AC	R89322;	
DE	13-APR-1996	(first entry)
DE	Rape seed beta-ketoacyl-ACP-ketoreductase.	
KW	Rape; seed; beta-ketoacyl-ACP-ketoreductase; beta-ketoreductase;	
KW	plasmid pJR510.1; CDNA library; embryo; Escherichia coli; vector;	
KW	plastid; stroma; transit peptide; cassette; anisense; oilseed;	
KW	transgenic plant; crop improvement; lipid; metabolic engineering;	
KW	polymer; rapeseed oil.	
OS	Brassica napus.	
FT	Key	Location/Qualifiers
FT	Note=	1..56
FT	/note="Transit peptide"	
PN	W09602652-A2.	
PD	01-FEB-1996.	
PF	17-JUL-1995; G01679.	
PR	20-JUL-1994; GB-014622.	
PA	(ZENEA) ZENEA LTD.	
PI	Chase D. Elborough K. Pentem PA, Siabas AF, White A;	

raised plants with enhanced oil content, diversion of metabolism to alternative
 low or modified oil content, starch, protein or engineered polymers, or
 storage compounds, e.g., starch, protein or engineered polymers, or
 production of plants with enhanced oil content. The DNA may also
 be used as a probe to obtain similar genes from other plants. The
 transit peptide may be used to direct other proteins to seed
 plastids.
 Sequence 315 AA;
 SQ

QY 69 SDEVINGPEQICKIV-GNIDVYHSIAFANNEDLRGFSREGFLAADISSULTI 127
Db 190 qa-atkinmkkkgrliniasvvgllngqanvaaaakagvigskaatargasrn.inv 248
QY 128 VAHEAKKLMP-EGGSIVATYLGGEFAVQNYNMVGAKASLEANNKYIALDLGPDNRVN 180
Db 249 vvcpgqfiasdmak-lge-d-mekkilgtiplrgyggpdedvagiiefialspaasyitgq 305

QY 187 AISAGPIRT-LSAKGVGGFNTILKEIERAPLRNVDQVEVGKTAAYL-LSDSLSSGVTE 244
 Db 306 aftidggi 313
 QY 245 NIHVDGSF 252

RESULT 12

ID R10974 standard; Protein; 246 AA.
 AC R10974;
 DT 17-APR-1991 (first entry)
 DE Acetoacetyl-CoA reductase enzyme.
 KW Polyester biopolymers; polyhydroxybutyrate; polyhydroxyalkanoate;
 KW beta-ketothiolase; acetoacetyl-CoA reductase.
 OS Alkaligenes eutrophus.
 PN W09100917-A.
 PD 24-JAN-1991.
 PF 10-JUL-1990; U03851.
 PR 10-JUL-1989; US-378155.
 (MASI) MASSACHUSETTS INST TECH.
 Peoples OP, Sinskey AJ;
 WPI; 91-051341/07.
 DR N-PSDB; Q10501.
 PT Construction and modification of polyester bio:polymers - by
 PT introduction of poly-hydroxy-butyrate and -alkanoate genes into
 PT bacteria or plants
 PS Disclosure; fig 3; 64pp; English.
 CC This Alkaligenes eutrophus acetyl-CoA reductase is an enzyme
 CC which is essential to the biosynthesis of polyhydroxyalkanoate
 CC (PHA). The gene encoding this is contained in plasmid clone,
 CC pAet3, downstream from the thiolase gene. The use of recombinant
 CC methods for producing such enzymes, required for polyester bio-
 CC polymer synthesis, allows for the control and modification of the
 CC synthesis process.
 CC See also Q10499-500 and Q10502-03.
 SQ Sequence 246 AA;

Query Match 8.0%; Score 140; DB 2; Length 246;
 Best Local Similarity 22.2%; Pred. NO. 9.35e-02;
 Matches 41; Conservative 59; Mismatches 77; Indels 8; Gaps 7;
 Db 66 dstktafdkvsksegevdvlinnagitrddvfr-kmtrawda-vidntlts--lfnvtk 121
 QY 71 EEVINGFEQIGKDVGNIDGVYHSIAFANMEDLRGRFSETRGFLAQDISSYSLTIVAH 130
 Db 122 qvidgmadrgrivnissvngkqgfgtqntystakaglhgftmaladevatkgtvntv 181
 131 EAKKLMPG-G-GSIVATTYLGGEFAVQNYVMGVAKASLEANKVYLALDLPDNRVNAI 188
 182 spgyiatdmvkairq-d-vldkivatipvkriglpesiasicawlseesgftgadfs 239
 QY 189 SAGPIRTLSAKGVGGFNTILKEIERAPLRNVDQVEVGKTAAYL-LSDSLSSGVTE 248
 Db 240 nggllh 244
 QY 249 DSGFH 253

RESULT 13

ID R32192 standard; Protein; 246 AA.
 AC R32192;
 DT 30-MAY-1993 (first entry)
 DE Sequence encoded by the acetoacetyl-CoA reductase (phbB) gene of the
 DE polyhydroxybutyrate (PHB) operon
 KW Operon; polyhydroxyalkanoate; acetoacetyl-CoA reductase.
 OS Alkaligenes eutrophus.
 PN W09302187-A.
 PD 04-FEB-1993.
 PF 13-JUL-1992; U05786.
 PR 19-JUL-1991; US-732243.
 PA (UYMA-) UNIV MADISON JAMES
 PA (UNMS) UNIV MICHIGAN STATE.

PI Dennis DE, Poirier Y, Somerville CR;
 DR WPI; 93-058785/07.
 DR N-PSDB; Q36660.
 PT Transgenic plants producing poly:hydroxy-alkanoate polymer(s) -
 PT obt. by transformation with DNA encoding 3-ketothiolase,
 PT acetoacetyl-CoA reductase and PHA synthase
 PS Disclosure; Fig 2; 70pp; English.
 CC The nucleotide sequence of the PHB operon was obtained from Janes, B.
 CC Hollar, J. and Dennis, D. in Dawes, E.A. (ed.) Novel Biodegradable
 CC Polymers, Kluwer Academic Publishers, 175-190 (1990). It contains
 CC the genes from PHB synthase, 3-ketothiolase and acetoacetyl-CoA
 CC reductase. The inventors claim a transgenic plant material contg.
 CC foreign DNA encoding a peptide which exhibits 3-ketothiolase activity,
 CC pref. where the DNA is an open reading from between nucleotides
 CC 2896-3877 (phb A gene), 842-2611 (phb C gene) or 3952-4692 (phb B
 CC gene) of the Alkaligenes eutrophus PHB operon.
 SQ Sequence 246 AA;

Query Match 8.0%; Score 140; DB 6; Length 246;
 Best Local Similarity 22.2%; Pred. NO. 9.35e-02;
 Matches 41; Conservative 59; Mismatches 77; Indels 8; Gaps 7;
 Db 66 dstktafdkvsksegevdvlinnagitrddvfr-kmtrawda-vidntlts--lfnvtk 121
 QY 71 EEVINGFEQIGKDVGNIDGVYHSIAFANMEDLRGRFSETRGFLAQDISSYSLTIVAH 130
 Db 122 qvidgmadrgrivnissvngkqgfgtqntystakaglhgftmaladevatkgtvntv 181
 131 EAKKLMPG-G-GSIVATTYLGGEFAVQNYVMGVAKASLEANKVYLALDLPDNRVNAI 188
 182 spgyiatdmvkairq-d-vldkivatipvkriglpesiasicawlseesgftgadfs 239
 QY 189 SAGPIRTLSAKGVGGFNTILKEIERAPLRNVDQVEVGKTAAYL-LSDSLSSGVTE 248
 Db 240 nggllh 244
 QY 249 DSGFH 253

RESULT 14

ID P94157 standard; protein; 246 AA.
 AC P94157;
 DT 12-JUN-1990 (first entry)
 DE Acetyl-CoA reductase.
 KW Polyester biopolymers; acetyl-CoA reductase; polyhydroxybutyrate; ss.
 OS Alkaligenes eutrophus H16.
 PN W08900202-A.
 PD 12-JAN-1989.
 PF 27-JUN-1988; U02227.
 PR 29-JUN-1987; US-067695.
 PA (MASI) Massachusetts Institute of Technology.
 PI Peoples OP, Sinskey AJ;
 DR WPI; 89-039655/05.
 DR N-PSDB; N91209.
 PT Constructing new polyester biopolymers - using genes encoding
 PT beta-ketothiolase(s), acetoacetyl-CoA reductase(s) and
 PT polyhydroxybutyrate synthetase(s).
 PS Disclosure; p; English.
 CC The enzyme is used to study the polyhydroxybutyrate (PHB) biosynthetic
 CC pathway. It catalyses the reduction of acetoacetyl-CoA to form
 CC D(-)-beta-hydroxybutyryl-CoA, the substrate for PHB synthetase, and can
 CC be used to control biopolymer syntheses and produce new biopolymers.
 CC The sequence was deduced from the DNA sequence.
 CC See also P94154-P94156.
 SQ Sequence 246 AA;

Query Match 8.0%; Score 140; DB 1; Length 246;
 Best Local Similarity 22.2%; Pred. NO. 9.35e-02;
 Matches 41; Conservative 59; Mismatches 77; Indels 8; Gaps 7;
 Db 66 dstktafdkvsksegevdvlinnagitrddvfr-kmtrawda-vidntlts--lfnvtk 121
 QY 71 EEVINGFEQIGKDVGNIDGVYHSIAFANMEDLRGRFSETRGFLAQDISSYSLTIVAH 130

Fri Dec 5 12:31:27 1997

Db 122 qvidgmadrqgrvinnssvngkqgfgtgnystakaglhgftmalagaevatkgtvntv 181
QY 131 EAKKLMEPE-G-GSIVATTYLGGEFAVQNYNVMGVAKASLEAVNKYLALDLPDNIIRVNAI 188
Db 182 spgyiatdmvkairq-d-vldkivatipvkrllgpeiasicawlseesgfgstgdfsl 239
QY 189 SAGPITLSAKGVGGFNTILKEIERAPLKRNVDOQEVGTAAYLLSDUSSGVTGENIHV 248
Db 240 nggllh 244
QY 249 DSGFPH 253

RESULT 15
ID R71324 standard; Protein; 329 AA.
AC R71324;
DT 21-OCT-1995 (first entry)
DE Acetyl-CoA-reductase.
poly-beta-hydroxyalkanoate; transgenic plant;
biodegradable thermoplastic.
OS Alkaligenes eutrophus.
PN W09505472-A.
PD 23-FEB-1995.
PF 17-AUG-1994; U09265.
PR 17-AUG-1993; US-108193.
PR 06-JUN-1994; US-254357.
PA (UNMS) UNIV MICHIGAN STATE.
PI Nawrath C. Poirier Y, Somerville CR;
DR WPI: 95-098770/13.
DR N-PSDB; Q85642.
PT Transgenic plant material with plastid(s) contg. the enzymes for
PT synthesis of poly(hydroxyalkanoate(s)) - express
PT poly(hydroxybutyrate and have good growth and seed formation.
PS Claim 2; Page 60-61; 88pp; English.
CC The acetyl-CoA-reductase gene (phbB) from A. eutrophus is cloned
CC under the control of an Arabidopsis thaliana seed storage protein
CC promoter for plastid tissue-specific gene expression in a
CC transgenic plant. When expressed with the 3-ketothiolase
CC (phbA) and poly-beta-hydroxyalkanoate-synthase (phbC) genes, a
CC poly-beta-hydroxyalkanoate (PHA), specifically poly-beta-
CC hydroxybutyrate (PHB), is expressed in the transgenic plant
CC (preferably a Brassica e.g. rape). PHB and related PHAs are
CC biodegradable thermoplastics with many useful applications.
SQ Sequence 329 AA;

Query Match 8.0%; Score 140; DB 13; Length 329;
1st Local Similarity 22.2%; Pred. No. 9.35e-02;
Matches 41; Conservative 59; Mismatches 77; Indels 8; Gaps 7;
Db 149 dsktafdkvvksegevdvlinnagitrddvfr-kmtradwda-vidtnlts--lfnvtk 204
QY 71 EEVINGFEQIGKDVGNIDGYHSHAFANMEDLRGFRSETSGREGFLLAQDISSVSLTIVAH 130
Db 205 qvidgmadrqgrvinnssvngkqgfgtgnystakaglhgftmalagaevatkgtvntv 264
QY 131 EAKKLMEPE-G-GSIVATTYLGGEFAVQNYNVMGVAKASLEAVNKYLALDLPDNIIRVNAI 188
Db 265 spgyiatdmvkairq-d-vldkivatipvkrllgpeiasicawlseesgfgstgdfsl 322
QY 189 SAGPITLSAKGVGGFNTILKEIERAPLKRNVDOQEVGTAAYLLSDUSSGVTGENIHV 248
Db 323 nggllh 327
QY 249 DSGFPH 253

Search completed: Thu Dec 4 16:01:20 1997
Job time : 41 secs.

WIRE (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
on: Thu Dec 4 16:10:30 1997; MasPar time 4.09 Seconds
Molecular output not generated. 307.588 Million cell updates/sec

Title: >US-08-790-043A-1
Description: (1-256) from US08790043A.pep
Perfect Score: 1753
Sequence: 1 MLNLENTKYVIMGIANKRSI.....LSSGVGTGENIHVDSGFHAIK 256

Scoring table: PAM 150
Gap 11

Searched: 55487 seqs, 4918616 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:back1 2:51 3:52 4:53 5:54 6:55 7:56 8:PT90 9:PT91
10:PT92 11:PT93 12:PT94 13:PT95 14:PT96

Statistics: Mean 29.990; Variance 162.121; scale 0.185

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Hit No.	Score	Query Match	Length	ID	Description	Pred. No.
1	142	8.1	246	1	5229279-7 Patent No. 5229279.	1.30e+02
2	108	6.2	273	1	5512669-4 Patent No. 5512669.	2.15e+00
3	108	6.2	273	1	5512669-4 Patent No. 5512669.	2.15e+00
4	100	5.7	418	11	PCT-US93-0 Sequence 3, Applicatio	6.79e+00
5	100	5.7	418	5	US-08-121- Sequence 3, Applicatio	6.79e+00
6	100	5.7	764	13	PCT-US95-1 Sequence 4, Applicatio	6.79e+00
7	100	5.7	764	13	US-08-375- Sequence 4, Applicatio	6.79e+00
8	100	5.7	1089	13	PCT-US95-1 Sequence 2, Applicatio	6.79e+00
9	100	5.7	1089	7	US-08-375- Sequence 2, Applicatio	6.79e+00
10	99	5.6	394	7	US-08-002- Sequence 9, Applicatio	7.82e+00
11	99	5.6	394	7	US-08-002- Sequence 11, Applicatio	7.82e+00
12	99	5.6	394	7	US-08-002- Sequence 11, Applicatio	7.82e+00
13	97	5.5	647	7	US-08-218- Sequence 6, Applicatio	7.82e+00
14	97	5.5	647	7	US-08-218- Sequence 6, Applicatio	7.82e+00
15	91	5.2	201	11	PCT-US93-1 Sequence 6, Applicatio	2.40e+01
16	92	5.2	263	5	US-07-721- Sequence 51, Applicati	2.09e+01
17	92	5.2	288	5	US-07-721- Sequence 52, Applicati	2.09e+01
18	91	5.2	453	4	US-08-089- Sequence 2, Applicatio	2.40e+01
19	91	5.2	453	6	US-08-421- Sequence 2, Applicatio	2.40e+01
20	92	5.2	582	6	US-08-261- Sequence 2, Applicatio	2.09e+01
21	89	5.1	444	1	5310667-1 Patent No. 5310667.	3.16e+01
22	90	5.1	3033	5	US-07-925- Sequence 8, Applicatio	2.76e+01

23	90	5.1	3033	5	US-07-925- Sequence 9, Applicatio	2.76e+01
24	88	5.0	3033	5	US-07-925- Sequence 5, Applicatio	3.63e+01
25	86	4.9	741	7	US-08-277- Sequence 4, Applicatio	4.77e+01
26	84	4.8	263	1	5229279-4 Patent No. 5229279.	6.25e+01
27	84	4.8	444	7	US-08-476- Sequence 55, Applicati	6.25e+01
28	84	4.8	444	7	US-08-306- Sequence 55, Applicati	6.25e+01
29	84	4.8	446	1	5310667-4 Patent No. 5310667.	6.25e+01
30	84	4.8	577	7	US-08-484- Sequence 24, Applicati	6.25e+01
31	84	4.8	577	6	US-08-484- Sequence 24, Applicati	6.25e+01
32	85	4.8	582	6	US-08-261- Sequence 4, Applicatio	5.46e+01
33	85	4.8	582	6	US-08-261- Sequence 6, Applicatio	5.46e+01
34	85	4.8	582	6	US-08-261- Sequence 8, Applicatio	5.46e+01
35	82	4.7	516	1	5188642-4 Patent No. 5188642.	8.18e+01
36	83	4.7	865	4	US-07-803- Sequence 13, Applicati	7.15e+01
37	82	4.7	1171	7	US-08-445- Sequence 1, Applicatio	8.18e+01
38	81	4.6	240	13	PCT-US95-0 Sequence 82, Applicati	9.34e+01
39	81	4.6	240	13	PCT-US95-0 Sequence 82, Applicati	9.34e+01
40	80	4.6	304	12	PCT-US94-0 Sequence 55, Applicati	1.07e+02
41	81	4.6	778	11	PCT-US93-0 Sequence 3, Applicatio	9.34e+01
42	81	4.6	778	11	PCT-US93-0 Sequence 4, Applicatio	9.34e+01
43	81	4.6	955	12	PCT-US94-0 Sequence 1, Applicatio	9.34e+01
44	81	4.6	955	5	US-08-006- Sequence 1, Applicatio	9.34e+01
45	81	4.6	1513	11	PCT-US93-0 Sequence 2, Applicatio	9.34e+01

ALIGNMENTS

RESULT 1
ID 5229279-7 STANDARD; PRT; 266 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Patent No. 5229279.
XX
CC Patent No. 5229279
CC APPLICANT: PEOPLES, OLIVER P.; SINSKEY, ANTHONY J.
CC TITLE OF INVENTION: METHOD FOR PRODUCING NOVEL POLYESTER BIOPOLYMERS
CC NUMBER OF SEQUENCES: 8
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/556,535
CC FILING DATE: 13-AUG-1990
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 67,695
CC FILING DATE: 29-AUG-1987
CC SEQ ID NO:7:
CC LENGTH: 246
CC SEQUENCE 266 AA; 28572 MW; 399914 CN;

Query Match 8.1%; Score 142; DB 1; Length 246;
Best Local Similarity 22.6%; Pred. No. 1.30e+02;
Matches 42; Conservative 60; Mismatches 74; Indels 10; Gaps 9;
Db 66 DSTKTAFAKVKSEGEVDVLINAGITRDVVR-KMTRADWDA-VIDFNLTSL-LENVTK 121
QY 71 EVINGFQIKGKDVNDGVVHSTAFANNEDLRGRFSREGFLAODISSYSLTIVAH 130
Db 122 QVIDGMADRGWRIYNISSVNGQKQFGQOTNYSTAKAGLHGFTMALAQEVATKGVTVNTV 181
QY 131 EAKKLMPG-G-GSIVATYVLGGEFAVQVNYNGVAKASLEANKVYALDLDPDIRVNAI 188

Db 182 SPGYTATDMVKAIRO-D-VLDKIYATIPVKRLGLPQ-BIASICAWLSSESGFTGADFS 238
QY 189 SAGPRTISAKGVGGFNTILKEIERAPLKR-NVDQVEVGKTAAYLLSLSGVTGENIH 247
Db 239 LNGGLH 244
QY 248 VDSGFH 253

RESULT 2

```

ID 5512669-4 STANDARD; PRT; 296 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Patent No. 5512669.
XX
CC Patent No. 5512669
CC APPLICANT: PEOPLES, OLIVER P.; SINSKEY, ANTHONY J.
CC TITLE OF INVENTION: GENE ENCODING BACTERIAL ACETOACETYL-COA
CC REDUCTASE
CC
CC NUMBER OF SEQUENCES: 7
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/297,667
CC FILING DATE: 29-AUG-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 124,570
CC FILING DATE: 20-SEP-1993
CC APPLICATION NUMBER: 944,488
CC FILING DATE: 03-NOV-1992
CC APPLICATION NUMBER: 566,535
CC FILING DATE: 13-AUG-1990
CC APPLICATION NUMBER: 67,695
CC FILING DATE: 29-JUN-1987
CC SEQ ID NO: 4:
CC LENGTH: 273
CC SEQUENCE 296 AA; 31345 MW; 472090 CN;

Query Match 6.2%; Score 108; DB 1; Length 273;
Best Local Similarity 27.4%; Pred. No. 2.15e+00;
Matches 31; Conservative 29; Mismatches 50; Indels 3; Gaps 2;

Db 159 GRVNISSINGKQGMQYNSAAKAGDLGFTKALAOEGAAGITVNAICPGYIGTEMVR 218
QY 140 GSIATYTLGGFAVQNVNMGVAKASLEANKYKALDGLGPDNIRVNAISAGPRTLSAK 199
Db 219 AIP--EKVLNERIIPQIPVGRIGDEPEIARIIVVFLASDEAGFTGTISANGG 269
QY 200 GVGGFNTILKE-IEERAPLKRNVQVEVGKTAAYLLSLDSSGVTGENIHVDSG 251

RESULT 3
ID 5512669-4 STANDARD; PRT; 296 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Patent No. 5512669
CC APPLICANT: PEOPLES, OLIVER P.; SINSKEY, ANTHONY J.
CC TITLE OF INVENTION: GENE ENCODING BACTERIAL ACETOACETYL-COA
CC REDUCTASE
CC
CC NUMBER OF SEQUENCES: 7
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/297,667
CC FILING DATE: 29-AUG-1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 124,570
CC FILING DATE: 20-SEP-1993
CC APPLICATION NUMBER: 944,488
CC FILING DATE: 03-NOV-1992
CC APPLICATION NUMBER: 566,535
CC FILING DATE: 13-AUG-1990
CC APPLICATION NUMBER: 67,695
CC FILING DATE: 29-JUN-1987
CC SEQ ID NO: 4:
CC LENGTH: 273
CC SEQUENCE 296 AA; 31345 MW; 472090 CN;

Query Match 6.2%; Score 108; DB 1; Length 273;
Best Local Similarity 27.4%; Pred. No. 2.15e+00;
Matches 31; Conservative 29; Mismatches 50; Indels 3; Gaps 2;

Db 159 GRVNISSINGKQGMQYNSAAKAGDLGFTKALAOEGAAGITVNAICPGYIGTEMVR 218
QY 140 GSIATYTLGGFAVQNVNMGVAKASLEANKYKALDGLGPDNIRVNAISAGPRTLSAK 199
Db 219 AIP--EKVLNERIIPQIPVGRIGDEPEIARIIVVFLASDEAGFTGTISANGG 269
QY 200 GVGGFNTILKE-IEERAPLKRNVQVEVGKTAAYLLSLDSSGVTGENIHVDSG 251

RESULT 4
ID PCT-US93-08322-3 STANDARD; PRT; 418 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 3, Application PC/TUS9308322.
XX
CC Sequence 3, Application PC/TUS9308322
CC GENERAL INFORMATION:
CC APPLICANT: Sager, Ruth
CC TITLE OF INVENTION: MASPIN, A NOVEL SERPIN WITH TUMOR SUPPRESSING ACT
CC IVITY
CC NUMBER OF SEQUENCES: 8
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson
CC STREET: 225 Franklin Street
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: U.S.A.
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
CC COMPUTER: IBM PS/2 Model 502 or 55SX
CC OPERATING SYSTEM: MS-DOS (Version 5.0)
CC SOFTWARE: WordPerfect (Version 5.1)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/08322
CC FILING DATE:
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/938,823
CC FILING DATE: 09/01/92
CC APPLICATION NUMBER: 07/844,296
CC FILING DATE: 02/28/92
CC APPLICATION NUMBER: 07/662,216
CC FILING DATE: 02/28/91
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fraser, Janis K.
CC REGISTRATION NUMBER: 34,819
CC REFERENCE/DOCKET NUMBER: 00530/072001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 542-5070
CC TELEFAX: (617) 542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 418
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC SEQUENCE 418 AA; 46736 MW; 945384 CN;

Query Match 5.7%; Score 100; DB 11; Length 418;
Best Local Similarity 25.4%; Pred. No. 6.79e+00;
Matches 16; Conservative 21; Mismatches 30; Indels 2; Gaps 2;

Db 100 ILEGLNFNLT-EIPEAQIHEGFQELLRLTNQPSQL-QLTGTGNGFLSGLKLVDFLED 157
QY 27 VLDQLGAKLVFTYRKERSRKELEKLEQLNQPEAHLYQIDVQSDDEVINGFEQIGKDVGN 86
```

	RESULT	6	PCT-US95-16930-4	STANDARD;	PRT;	764 AA.
XX	ID	AC	xxxxxx			
XX	DT	DT	01-JAN-1900			
XX	DE		Sequence 4, Application PC/TUS9516930.			
sequence 4, Application PC/TUS9516930 GENERAL INFORMATION: APPLICANT: UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE TITLE OF INVENTION: PRODUCTION IN THE ABSENCE OF TITLE OF INVENTION: NONSENSE-MEDIATED mRNA DECAY TITLE OF INVENTION: FUNCTION NUMBER OF SEQUENCES: 6 CORRESPONDENCE ADDRESS: ADDRESSEE: Fish & Richardson STREET: 225 Franklin Street Suite 3100 CITY: Boston STATE: MA COUNTRY: USA ZIP: 02110-2804 COMPUTER READABLE FORM: MEDIA TYPE: Floppy disk COMPUTER: IBM PC compatible OPERATING SYSTEM: PC-DOS/MS-DOS SOFTWARE: PatentIn Release #1.0, Version #1.30 CURRENT APPLICATION DATA: APPLICATION NUMBER: PCT/US95/16930 FILING DATE: 27-DEC-1995 PRIOR APPLICATION DATA: APPLICATION NUMBER: US 08/375,300 FILING DATE: 20-JAN-1995 ATTORNEY/AGENT INFORMATION: NAME: Fasse, J. Peter REGISTRATION NUMBER: 32,983 REFERENCE/DOCKET NUMBER: 04020/0465W01 TELECOMMUNICATION INFORMATION: TELEPHONE: (617)542-5070 TELEFAX: (617)542-8906 TELEX: 200154 INFORMATION FOR SEQ ID NO.: 4: SEQUENCE CHARACTERISTICS: LENGTH: 764 amino acids TYPE: amino acid STRANDNESS: single TOPOLOGY: linear MOLECULE TYPE: protein SEQUENCE 764 AA; 89464 MW; 2913896 CN; Query Match 5.7%; Score 100; DB 13; Length 764 Best Local Similarity 29.1% ; Pred.No.6,79e+00; Matches 25; Conservative 25; Mismatches 30; Indels						
D b	307	LTKVGLGYSYRDFIVRCIDQVLNTERGLE-IINDYGQNMRHSNNRYLTETII	:::: : : : : :::: : : : :	:	:	:
Q y	24	VAKVLDQL-GAKLVFTYR-KERSRKELKULLELNQPEAHL Y Q I -DVQSDEV	: : : : : : : :		:	:
D b	365	KSDV-LDDTYHIRFGHHNPFP 389	: : : : : :		:	:
Q y	81	GKDVGNDGVYHSIAFNMEDLRGF 106	: : : : : :		:	:
RESULT	7	US-08-375-300-4	STANDARD;	PRT;	764 AA.	
XX	ID	AC	xxxxxx			
XX	DT	DT	01-JAN-1900			

```

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Fish & Richardson
CC STREET: 225 Franklin Street Suite 3100
CC CITY: Boston
CC STATE: MA
CC COUNTRY: USA
CC ZIP: 02110-2804
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/16930
CC FILING DATE: 27-DEC-1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/375,300
CC FILING DATE: 20-JAN-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fasse, J. Peter
CC REGISTRATION NUMBER: 32,983
CC REFERENCE/DOCKET NUMBER: 04020/046WO1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617)542-5070
CC TELEFAX: (617)542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 2:
CC Best Local Similarity 29.1%; Score 100; DB 13; Length 1089;
CC Matches 25; Conservative 25; Mismatches 30; Indels 6; Gaps 6;
CC
CC Db 632 LTKVLGLGYRREDFRCIDOVLENIERCLE-INDYQGNMHRISNRYLTFIFN-FEMI 689
CC QY 24 VAKVLDOL-CAKLVFFYR-KERSRKELEKLELNQPEAHLQYI-DVQSDSEVINGFEQI 80
CC
CC Db 590 KSDV-LLDTYHIRFGHINNQPWF 714
CC QY 81 GKDVGNIDGVYHSAFANMEDLRGF 106
CC
CC RESULT 9
CC ID US-08-375-300-2 STANDARD; PRT; 1089 AA.
CC AC xxxxxx
CC XX
CC XX 01-JAN-1900
CC DT
CC XX
CC XX Sequence 2, Application US/08375300.
CC CC Sequence 2, Application US/08375300
CC CC Patent No. 5679566
CC CC GENERAL INFORMATION:
CC CC APPLICANT: Feng, He
CC CC APPLICANT: Jacobson, Allan S.
CC CC TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN
CC CC TITLE OF INVENTION: THE ABSENCE OF NONSENSE-MEDIATED MRNA DECAY FUNC
CC CC ION
CC CC NUMBER OF SEQUENCES: 6
CC CC CORRESPONDENCE ADDRESS:
CC CC ADDRESSEE: Fish & Richardson
CC CC STREET: 225 Franklin Street Suite 3100
CC CC CITY: Boston
CC CC STATE: MA
CC CC COUNTRY: USA
CC CC ZIP: 02110-2804

```

CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/375,300
CC FILING DATE: 20-JAN-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fasse, J. P.
CC REGISTRATION NUMBER: 32,983
CC REFERENCE/DOCKET NUMBER: 04020/046001
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617)542-5070
CC TELEFAX: (617)542-8906
CC TELEX: 200154
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1089 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: not relevant
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE: 1089 AA; 126746 MW; 5973553 CN;
SQ
Query Match 5.7%; Score 100; DB 7; Length 1089;
Best Local Similarity 29.1%; Pred. No. 6.79e+00;
Matches 25; Conservative 25; Mismatches 30; Indels 6; Gaps 6;
Db 632 LTKVLGLGYRFRDTCIDVLENIERGLE-INDYQNMHRISNRYLTETFN-FEMI 689
QY 24 VAKVLDGL-CAKLVFYFR-KERSRKELEKLEQLNQPEAHLQI-DVQSDSEEVINGFEQI 80
Db 690 KSDV-LLDTIYHIRFGHINNPNPF 714
QY 81 GKDVGNIDGVYHSTAFANMEDLRGRF 106
RESULT 10
ID US-08-002-202-9 STANDARD; PRT; 394 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 9, Application US/08002202.
XX
CC Sequence 9, Application US/08002202
CC Patent No. 5604201
CC GENERAL INFORMATION:
CC APPLICANT: Thomas, Garry
CC APPLICANT: Anderson, Eric D
CC APPLICANT: Thomas, Laurel
CC APPLICANT: Hayflick, Joel S
CC TITLE OF INVENTION: Methods and Reagents for Inhibiting
CC TITLE OF INVENTION: Furin Endoprotease
CC NUMBER OF SEQUENCES: 21
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Allegretti and Witcoff, Ltd.
CC STREET: 10 South Wacker Drive, Suite 3000
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60606
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/002,202
CC FILING DATE: 08-JAN-1993
CC CLASSIFICATION: 514

CC ATTORNEY/AGENT INFORMATION:
CC NAME: NO. 5604201nan, Kevin E
CC REGISTRATION NUMBER: 35,30003
CC REFERENCE/DOCKET NUMBER: 92,448
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 312-715-1000
CC TELEFAX: 312-715-1234
CC TELEX: 910-221-5317
CC INFORMATION FOR SEQ ID NO: 9:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 394 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC FEATURE:
CC NAME/KEY: Protein
CC LOCATION: 1..394
CC OTHER INFORMATION: /label= Variant
CC OTHER INFORMATION: /note= "This amino acid sequence is the amino ac
id
CC OTHER INFORMATION: sequence of the modified alpha-1-antitrypsin
CC OTHER INFORMATION: protein, alpha-1-antitrypsin Portland;"
SQ SEQUENCE 394 AA; 44432 MW; 838791 CN;
Query Match 5.6%; Score 99; DB 7; Length 394;
Best Local Similarity 25.4%; Pred. No. 7.82e+00;
Matches 18; Conservative 20; Mismatches 31; Indels 2; Gaps 2;
Db 76 ILEGLNLFNL-ETPEAQIHGFEQLLRLNQPSQL-QLTTGNGLFLSQGLKLVDFKFLD 133
QY 27 VLDQLGAKLVFYRKRSRKELEKLEQLNQPEAHLQI-DVQSDSEEVINGFEQIGKDVGN 86
Db 134 VKKLYHSEAPT 144
QY 87 IDGVYHSTAF 97
RESULT 11
ID US-08-002-202-11 STANDARD; PRT; 394 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 11, Application US/08002202.
XX
CC Sequence 11, Application US/08002202
CC Patent No. 5604201
CC GENERAL INFORMATION:
CC APPLICANT: Thomas, Garry
CC APPLICANT: Anderson, Eric D
CC APPLICANT: Thomas, Laurel
CC APPLICANT: Hayflick, Joel S
CC TITLE OF INVENTION: Methods and Reagents for Inhibiting
CC TITLE OF INVENTION: Furin Endoprotease
CC NUMBER OF SEQUENCES: 21
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Allegretti and Witcoff, Ltd.
CC STREET: 10 South Wacker Drive, Suite 3000
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60606
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/002,202
CC FILING DATE: 08-JAN-1993
CC CLASSIFICATION: 514
CC ATTORNEY/AGENT INFORMATION:

CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 647 AA; 72413 MW; 2325170 CN;
Query Match 5.5%; Score 97; DB 7; Length 647;
Best Local Similarity 28.9%; Pred. No. 1.04e+01;
Matches 26; Conservative 23; Mismatches 34; Indels 7; Gaps 6;
Db 249 VGNAPYIVPDDPRSMFAETGIFRSPGHEDHLAGBEIASYS--IVGPANAK-VPHSA 305
QY 84 VGNIDGVY-HSTA-FANMEDLRGRFSETSREGF-LLAQDISSYSLTIVAHEAKKLMPGG 140
Db 306 S-SDTSLIAYSGIPSYSLSLTSTAK 334
QY 141 SIVATTYLGGEFAVQVNMVGVAKASLEAN 170
RESULT 14
ID US-08-292-945-6 STANDARD; PRT; 201 AA.
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 6, Application US/08292945.
XX
Sequence 6, Application US/08292945
Patent No. 5585478
GENERAL INFORMATION:
CC APPLICANT: Lim, Bing
CC APPLICANT: Lelias, Jean-Michel
CC APPLICANT: Adra, Chaker
CC TITLE OF INVENTION: D4 Gene and Methods of Use Thereof
CC NUMBER OF SEQUENCES: 8
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Kilpatrick & Cody
CC STREET: 1100 Peachtree Street, Suite 2800
CC CITY: Atlanta
CC STATE: Georgia
CC COUNTRY: United States
CC ZIP: 30309-4530
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/292,945
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/990,337
CC FILING DATE: December 10, 1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Pabst, Patrea L.
CC REGISTRATION NUMBER: 31,284
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (404) 815-6508
CC TELEFAX: (404) 815-6555
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 201 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal

CC ORIGINAL SOURCE:
CC ORGANISM: Murine
SQ SEQUENCE 201 AA; 22979 MW; 226152 CN;
Query Match 5.2%; Score 91; DB 6; Length 201;
Best Local Similarity 23.0%; Pred. No. 2.40e+01;
Matches 28; Conservative 27; Mismatches 63; Indels 4; Gaps 4;
Db 15 DDLDSKLNKYPPOKSLKLEQM-DKDDSLTK-YKKTLLGDVPVADPTVPNVTVTRL 72
QY 29 DDLGAKLVFTYRKERSRKEKLEQLNQPEAHLYQIDVQSDSEEVINGFEQIGKDVGNID 88
Db 73 LVCDSPAGPITMDLTGDLKAKKDTFVLKEGIE-YRVKINFKVKNKDIVS-GLKYVQHYR 130
QY 89 GYHSAIFANMEDLRGRFSETSREGFLLAQDISSYSLTIVAHEAKKLMPGGSIIVAT 148
Db 131 TG 132
QY 149 GG 150
RESULT 15
ID PCT-US93-12074-6 STANDARD; PRT; 201 AA.
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 6, Application PC/TUS9312074.
XX
Sequence 6, Application PC/TUS9312074
GENERAL INFORMATION:
CC APPLICANT:
CC TITLE OF INVENTION: D4 Gene and Methods of Use Thereof
CC NUMBER OF SEQUENCES: 8
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/12074
CC FILING DATE:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/990,337
CC FILING DATE: 10-DEC-1992
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 201 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: Murine
CC SEQUENCE 201 AA; 22979 MW; 226152 CN;
Query Match 5.2%; Score 91; DB 11; Length 201;
Best Local Similarity 23.0%; Pred. No. 2.40e+01;
Matches 28; Conservative 27; Mismatches 63; Indels 4; Gaps 4;
Db 15 DDLDSKLNKYPPOKSLKLEQM-DKDDSLTK-YKKTLLGDVPVADPTVPNVTVTRL 72
QY 29 DDLGAKLVFTYRKERSRKEKLEQLNQPEAHLYQIDVQSDSEEVINGFEQIGKDVGNID 88
Db 73 LVCDSPAGPITMDLTGDLKAKKDTFVLKEGIE-YRVKINFKVKNKDIVS-GLKYVQHYR 130
QY 89 GYHSAIFANMEDLRGRFSETSREGFLLAQDISSYSLTIVAHEAKKLMPGGSIIVAT 148
Db 131 TG 132

Fri Dec 5 12:31:28 1997

US-08-790-043A-1.rai

Page 8

Qy 149 GG 150

Search completed: Thu Dec 4 16:11:18 1997
Job time : 48 secs.

WILEY

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.

Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.

Distribution rights by IntelliGenetics, Inc.

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

on on: Thu Dec 4 16:01:38 1997; Maspar time 12.50 Seconds

Bular output not generated. 591.864 Million cell updates/sec

Title: >US-08-790-043A-1

Description: (1-256) from US087900043A.pap

Perfect Score: 1753

Sequence: 1 MLNLENTYVINGIAKRSI.....LSSGVTGENIHVDSGFHAIK 256

Scoring table: PAM 150

Gap 11

Searched: 91006 seqs, 2888923 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: pir51

1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3

8:unann4 9:unann5 10:unann6 11:unann7 12:unann8

13:unann9 14:unann10 15:unenc 16:unrev

Statistics: Mean 47.222; Variance 135.383; scale 0.349

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Alt No.	Score	Match	Query Length	ID	Description	Pred. No.
1	748	42.7	262	9	S48029 short-chain alcohol	2.41e-84
2	732	41.8	262	9	B43729 envM protein - Salm	4.58e-82
3	718	41.0	295	9	B41139 short chain alcohol	4.49e-80
4	365	20.8	385	16	enoyl-acyl carrier p	4.05e-31
5	229	13.1	249	5	B37761 bile acid 7-dehydrox	1.30e-13
6	224	12.8	244	5	JN0703 carbonyl reductase	5.31e-13
7	225	12.8	249	5	A31841 27K-2 protein (choli	4.01e-13
8	219	12.5	244	5	A28053 adipocyte p27 protei	2.15e-12
9	214	12.2	260	11	B48674 tropinone reductase	8.66e-12
10	212	12.1	249	5	A28212 27K bile acid 7-dehy	1.51e-11
11	197	11.2	245	1	C24706 nodulation protein n	9.25e-10
12	190	10.8	266	5	A42468 7alpha-hydroxysterol	6.14e-09
13	183	10.4	254	9	S56492 hypothetical protein	3.99e-08
14	182	10.4	254	9	S34724 probable oxidoreduct	5.20e-08
15	181	10.3	262	16	A48950 ariatoxin biosynthes	6.78e-08
16	179	10.2	255	5	A38527 7alpha-hydroxysterol	1.15e-07
17	171	9.8	256	8	A57149 gluconate 5-dehydrog	9.38e-07
18	171	9.8	320	5	S22450 3-oxoacyl-acyl-carr	9.38e-07
19	169	9.6	242	11	S39508 hypothetical protein	1.58e-06
20	169	9.6	261	9	S56475 hypothetical protein	1.58e-06
21	168	9.6	268	11	C48674 tropinone reductase	2.04e-06

22	167	9.5	246	1	DEKNG	acetoacetyl-CoA redu	2.65e-06
23	166	9.5	246	16	S58176	mithramycin polyketi	3.43e-06
24	166	9.5	278	11	S57351	NAD(+) dependent D-a	3.43e-06
25	164	9.4	244	1	B42147	3-oxoacyl-acyl-carr	5.73e-06
26	161	9.2	247	8	A55208	short-chain alcohol	1.23e-05
27	162	9.2	278	5	S01065	fixr protein - Brady	9.56e-06
28	159	9.1	295	5	S00729	sporulation protein	2.05e-05
29	157	9.0	283	12	S41412	tetrahydroxynaphthal	3.41e-05
30	156	8.9	114	10	S43610	hypothetical protein	4.38e-05
31	156	8.9	140	10	S43612	hypothetical protein	4.38e-05
32	155	8.8	273	11	A48674	tropinone reductase	5.64e-05
33	155	8.8	335	13	S53352	2,4-dienoyl-CoA redu	5.64e-05
34	153	8.7	267	10	B59850	hypothetical protein	9.31e-05
35	153	8.7	282	11	JC4041	D-arabinitol dehydro	9.31e-05
36	148	8.4	242	5	D64051	3-ketoacyl-acyl carr	3.23e-04
37	148	8.4	255	8	S39654	hypothetical protein	3.23e-04
38	147	8.4	255	10	S39737	hypothetical protein	4.13e-04
39	140	8.0	246	1	RDALAE	acetoacetyl-CoA redu	2.28e-03
40	140	8.0	289	5	S25286	probable dehydrogena	2.28e-03
41	139	7.9	249	5	S05398	granatidin-producing	2.90e-03
42	138	7.9	261	10	I39853	glucose dehydrogenas	3.68e-03
43	137	7.8	261	5	S35196	hypothetical protein	4.68e-03
44	135	7.7	261	10	I40224	glucose 1-dehydrogen	7.54e-03
45	135	7.7	271	5	A43744	N-acylmannosamine 1-	7.54e-03

ALIGNMENTS

RESULT 1
ENTRY S48029 #type complete
TITLE short-chain alcohol dehydrogenase homolog envM - Escherichia coli
ALTERNATE_NAMES enoyl-ACP reductase
ORGANISM #formal_name Escherichia coli
DATE 07-May-1995 #sequence_revision 21-Jul-1995 #text_change

ACCESSIONS S48029; A47681
REFERENCE 27-Oct-1995

#authors Kater, M.M.; Koningstein, G.M.; Nijkamp, H.J.J.; Stuitje, A.R.

#journal Plant Mol. Biol. (1994) 25:771-790

#title The use of a hybrid genetic system to study the functional relationship between prokaryotic and plant multi-enzyme fatty acid synthetase complexes.

#accession S48029

##status Preliminary

##molecule_type DNA

##residues 1-262 ##label KAT

##cross-references EMBL:X78733

REFERENCE A47681

#authors Bergler, H.; Hogenauer, G.; Turnowsky, F.

#journal J. Gen. Microbiol. (1992) 138:2093-2100

#title Sequences of the envM gene and of two mutated alleles in Escherichia coli.

#cross-references MUID:93123967

#accession A47681

##status Preliminary

##molecule_type DNA

##residues 1-262 ##label BER

##cross-references NCBI:121825; NCBI:121826

##note sequence extracted from NCBI backbone

#length 262 #molecular-weight 27864 #checksum 2286

Query Match 42.7%; Score 748; DB 9; Length 262;

Best Local Similarity 44.4%; Pred. No. 2.41e-84;

Matches 111; Conservative 65; Mismatches 71; Indels 3; Gaps 3;

Db 4 lsgkrilvtgvasikisvyaqamhregaelafyqndklkgvvefaaqlg-sdiyl- 61

QY 4 LENKTYVINGIAKRSIAGVAKVLDQIGAKLVFTYRKSRKELEKLEQNLQNPFAHLY 63

Db 62 qcdvaedsidtmfaelgkvwpkfdgfvhsigfapgdqldgdyvnavtregfkiahdlss 121

```
QY 64 QIDVQSDVEEVINGEQIGKDVGNIDGVIHSAFANMEDLGRFSET-SREGFLAQDISS 122
Db 122 ysfvamaacacrlmngpsalltsylgaeraipnynvmglakasleavrymanamgpeg 181
QY 123 YSLTIIVAHEAKKLMPEGGSVATYILGGFAVQNYVMGAKASLEAVRYMADLGLGPDN 182
Db 182 yrvnaisagpirtlaasgikdkfmlahceavtpirtvtiedvgnsaafclsdlsagis 241
QY 183 IRVNAISAGPRTLSAKGVGGFTILKEIERAPLRNVQVEVGTAAYLLDLSGGVT 242
Db 242 gevvhvdggf 251
QY 243 GENIHVDSGF 252

RESULT 2 B43729 #type complete
ENTRY envm protein - Salmonella typhimurium
TITLE #formal_name Salmonella typhimurium
ORGANISM 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change
DATE 23-Jun-1993
ACCESSIONS B43729
REFERENCE A43729
#authors Turnowsky, F.; Fuchs, K.; Jeschek, C.; Hoegenauer, G.
#journal J. Bacteriol. (1989) 171:6555-6565
#title envM genes of Salmonella typhimurium and Escherichia coli.
#accession B43729
#status preliminary
#molecule_type DNA
#residues 1-262 #label TUR
#cross-references GB:M31806
SUMMARY #length 262 #molecular-weight 27761 #checksum 2093

Query Match 41.8%; Score 732; DB 9; Length 262;
Best Local Similarity 43.8%; Pred. No. 4.58e-82;
Matches 110; Conservative 65; Mismatches 71; Indels 5; Gaps 5;

Db 4 lsgkrilvtgasklsaygiagahregaelafatyndklgrveefaaglgssiv-1- 61
QY 4 LENKTYVINGIAKRSAIFGAVKVDLGAKLVFTYRKERSRKELEKLEQLNQPEAHLY 63

Db 62 pcdvaedasidamfaelg-nvwpkfdgfvhsigfapgdqldgdyvnavtrregfkvhadis 120
QY 64 QIDVQSDVEEVINGEQIGKDV-GNIDGVYHSAFANMEDLGRFSET-SREGFLAQDIS 121
Db 121 ysfvamaacacrlmngpsalltsylgaeraipnynvmglakasleavrymanamgpeg 180
QY 122 YSLTIIVAHEAKKLMPEGGSVATYILGGFAVQNYVMGAKASLEAVRYMADLGLGPD 181
Db 181 gvrnaisagpirtlaasgikdkfmlahceavtpirtvtiedvgnsaafclsdlsagi 240
QY 182 NIRVNAISAGPRTLSAKGVGGFTILKEIERAPLRNVQVEVGTAAYLLDLSGGV 241
Db 241 sgevvhvdggf 251
QY 242 TGENIHVDSGF 252

RESULT 3 B64139 #type complete
ENTRY short chain alcohol dehydrogenase homolog (envM) homolog -
TITLE Haemophilus influenzae (strain Rd KW20)
ORGANISM #formal_name Haemophilus influenzae
DATE 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
10-May-1996
ACCESSIONS B64139
REFERENCE A64000
#authors Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.;
Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
Fitzhugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
```

```
M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
Science (1995) 269:496-512
#journal Whole-genome random sequencing and assembly of Haemophilus
#title influenzae Rd.
#accession B64139
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-295 #label TIGR
#cross-references GB:L42023; TIGR:HI1734
#note start codon was translated as Val
#note named as homolog to a protein from Escherichia coli
SUMMARY #length 295 #molecular-weight 31994 #checksum 5301

Query Match 41.0%; Score 718; DB 9; Length 295;
Best Local Similarity 43.2%; Pred. No. 4.49e-80;
Matches 108; Conservative 70; Mismatches 69; Indels 3; Gaps 3;

Db 37 ltgkrilvtglasnrslaygiaksmkeqgaelaftyndklgrveefakefg-sdivl- 94
QY 4 LENKTYVINGIAKRSAIFGAVKVDLGAKLVFTYRKERSRKELEKLEQLNQPEAHLY 63

Db 95 pldvatdesiqncfaelskrwdkfdgfihafapgdqldgdyvnaatregyriahdisa 154
QY 64 QIDVQSDVEEVINGEQIGKDVGNIDGVYHSAFANMEDLGRF-SETSREGFLAQDISS 122
Db 155 ysfvamaacacrlmngpsalltsylgaeraipnynvmglakasleavrymaadlgkeg 214
QY 123 YSLTIIVAHEAKKLMPEGGSVATYILGGFAVQNYVMGAKASLEAVRYMADLGLGPD 182
Db 215 irvnaisagpirtlaasgikdkfmlstfekttaalrtvtiedvgnsaafclsdlsag 274
QY 183 IRVNAISAGPRTLSAKGVGGFTILKEIERAPLRNVQVEVGTAAYLLDLSGGVT 242
Db 275 geivhvdagf 284
QY 243 GENIHVDSGF 252

RESULT 4 S17761 #type complete
ENTRY enoyl-acyl carrier protein reductase - rape
TITLE #formal_name Brassica napus #common_name rape
ORGANISM 22-Nov-1993 #sequence_revision 22-Nov-1993 #text_change
DATE 22-Nov-1993
ACCESSIONS S17761
REFERENCE S17761
#authors Kater, M.M.; Koningsstein, G.M.; Nijkamp, H.J.J.; Stuitje,
A.R.
#journal Plant Mol. Biol. (1991) 17:895-909
#title cDNA cloning and expression of Brassica napus enoyl-acyl
carrier protein reductase in Escherichia coli.
#cross-references MUID:92003699
#accession S17761
#status preliminary
#residues 1-385 #label KAT
SUMMARY #length 385 #molecular-weight 40486 #checksum 4099

Query Match 20.8%; Score 365; DB 16; Length 385;
Best Local Similarity 38.1%; Pred. No. 4.05e-31;
Matches 67; Conservative 54; Mismatches 49; Indels 6; Gaps 5;

Db 199 dfgsidilvhslna-ngepevskpilletstkgylaaasysfvsilsh-flpimppgga 255
QY 83 DVGNDGVYHSAFANMEDLGRFSET-SREGFLAQDISSYL-TIVAHEAKKLMPEGGS 141
Db 256 sisltyiaserilpgygggmssakaalesdtrvlfafagrknirvntisagplgsraak 315
QY 142 IVATTYLGGEFAVQNYVMGAKASLEAVRYMADLGLGPD-DNIRVNAISAGPRTLSAK 199
```

ACCESSIONS

JN0703; PNO569
JN0703
#authors Nakanishi, M.; Deyashiki, Y.; Nakayama, T.; Sato, K.; Hara,
A.

#journal Biochem. Biophys. Res. Commun. (1993) 194:1311-1316
#title Cloning and sequence analysis of a cDNA encoding tetrameric carbonyl reductase of pig lung.

#accession JN0703
##molecule_type mRNA
##residues 1-244 ##label NAK
#accession PNO569
##molecule_type protein
##residues 1-22;123-136;138-151;166-175;187-201;240-244 #label NA2

COMMENT
This enzyme is an NADPH-linked oxidoreductase that catalyzes the reduction of various carbonyl compounds to the corresponding alcohols.

CLASSIFICATION #superfamily ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
KEYWORDS NADP; oxido-reductase

FEATURE
8-180 #domain short-chain alcohol dehydrogenase homology
 #label SADH
 #length 244 ##molecular-weight 25986 #checksum 8918

SUMMARY
Query Match 12.8%; Score 224; DB 5; Length 244;
Best Local Similarity 24.2%; Pred. No. 5,31e-13;
Matches 44; Conservative 56; Mismatches 78; Indels 4; Gaps 3;

Dd 61 lgdwateralgvpgvdyllvnnaava-lmqpldttkkevdfdrsfvnlnrsvfgvsqiva 119
| : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 74 INGFEGIKGVGIDGYIHYSIAFANNEDLRGRFSREGLFIADISYSLTIVAHEAK 133
| : | : | : | : | : | : | : | : | : | : | : | : | :

Dd 120 rsmliervgpsivnvasmsvshvtypglaaystkgamtltksameigphkirvnsvp 179
| : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 134 KLMEPG-G--GSIVATTVLGEFAVQNVMVGAKASLEANKVKYLALDLGPONIRVNAlSA 190

Dd 180 tvvitamgrsvtsdpelarlkherpmrkfaevdvwnsilflldrasastgsisfida 239
| : | : | : | : | : | : | : | : | : | : | : | : | :
Qy 191 GPIRTLGAQGGVFNTTLKEIERAPLKRVNDQVEVGKTAAYLLSDLSGVGTGENIHWDS 250

Dd 240 gy 241
| :
Qy 251 GF 252

RESULT 7
ENTRY A31841 #type complete
TITLE 27K-2 protein (cholic acid-induced) - Eubacterium sp.
ALTERNATE_NAMES baiA2_27K protein
ORGANISM #formal_name Eubacterium sp.
DATE 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 23-Mar-1995

ACCESSIONS A31841; E37844
REFERENCE A31886
#authors White, W.B.; Franklund, C.V.; Coleman, J.P.; Hylemon, P.B.
#journal J. Bacteriol. (1988) 170:4555-4561
#title Evidence for a multigenic family involved in bile acid 7-dehydroxylation in Eubacterium sp. strain VPI 12708.

#cross-references UID:89008068
#accession A31841
##molecule_type DNA
##residues 1-249 ##label WHI

REFERENCE A37844
#authors Mallonee, D.H.; White, W.B.; Hylemon, P.B.
#journal J. Bacteriol. (1990) 172:7011-7019
#title Cloning and sequencing of a bile acid-inducible operon from Eubacterium sp. strain VPI 12708.

#cross-references UID:91072253
#accession E37844 preliminary
##status preliminary
##molecule_type DNA
##residues 1-249 ##label MAL
##cross-references GB:M36292

```

Db      240 gy 241
      1:
QY      251 GF 252

RESULT      9
ENTRY      B48674      #type complete
TITLE      tropinone reductase (EC 1.1.1.236) II - jimsonweed
ORGANISM   #formal_name detura stramonium #common_name jimsonweed,
            common thornapple
DATE       03-May-1994 #sequence_revision 03-May-1994 #text_change
            20-Mar-1996
ACCESSIONS B48674
REFERENCE  A48674
            Nakajima, K.; Hashimoto, T.; Yamada, Y.
            Proc. Natl. Acad. Sci. U.S.A. (1993) 90:9591-9595
            Two tropinone reductases with different stereospecificities
            are short-chain dehydrogenases evolved from a common
            ancestor.
#accession B48674
#status    preliminary
#molecule_type mrna
#residues  1-260 #label NAK
#cross-references GB:L20474
KEYWORDS   Oxidoreductase
SUMMARY    #length 260 #molecular-weight 28311 #checksum 3973

Query Match      12.2%; Score 214; DB 11; Length 260;
Best Local Similarity 24.5%; Pred. No. 8.66e-12;
Matches 63; Conservative 82; Mismatches 97; Indels 15; Gaps 15;

Db      6 nlegctaltvg-gs-rigivgyiveelasgas-vytcsrnqkelndcltqwrskgfkvea 62
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      3 NLEKNTVIMGIANKRSIAFGVAKVLDLGAKLVFTY-RKRSRKE-LEKLLLEQLNQPEA 60
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db      63 svcdlsrsrerqelmtvnhf-h-gklnilvnnagiviykeakd-ytvded-yslimsin 118
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      61 HLYQIDVQSD-EEVINGFEIGIKDVGNDGVHSIAFANNEDLGRFSETSGREGFLAOD 119
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db      119 feaahylsvlahpflk-aserngvfvissvgalavpyeavvgatkgamdqitrcrlafew 177
      :::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      120 I-SSYSTIIVAEAKKLMPGGSGIVATTVLGGEFAVQNTNVNGVAKASLEANKVLAIDL 178
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db      178 akdnirvngvgpvviatsvemtqdpckenlkliidrcalrrngpekelaaamvaficf 237
      :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      179 GPDNIRVNAISAGPIT-LSAKGVGGFN-TI-LKEIERAPLKRNVQVQVEGKTAAYLLS 235
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db      238 paasyvtgqilyvdggl 254
      :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      236 DLSSGVTGENIHVDSGF 252

RESULT      10
ENTRY      A28212      #type complete
TITLE      27K bile acid 7-dehydroxylation protein - Eubacterium sp.
ORGANISM   #formal_name Eubacterium sp.
DATE       30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change
            12-Apr-1995
ACCESSIONS A28212
REFERENCE  A28212
            Coleman, J.P.; White, W.B.; Lijewski, M.; Hylemon, P.B.
            J. Bacteriol. (1988) 170:2070-2077
            Nucleotide sequence and regulation of a gene involved in bile
            acid 7-dehydroxylation by Eubacterium sp. strain VPI 12708.
#cross-references MUID:88197993
#accession A28212
#molecule_type DNA
#residues  1-249 #label COL
#experimental_source strain VPI 12708
CLASSIFICATION #superfamily ribitol dehydrogenase; short-chain alcohol
                dehydrogenase homology
FEATURE
7-188      #domain short-chain alcohol dehydrogenase homology

```

SUMMARY #length 249 #molecular-weight 26747 #checksum 341

Query Match 12.1%; Score 212; DB 5; Length 249;
Best Local Similarity 23.0%; Pred. No. 1.51e-11;
Matches 59; Conservative 80; Mismatches 106; Indels 12; Gaps 12;

Db 1 mklvqdktiitg-gt-rigifaaakfiengakvsifgetqeevdtalaql-kelypee 57
Qy 2 LNL-ENKTYVINGANKRSIAFGVAKVLDQAKL-VFTYRKERSRKELEKLEQLNQPE 59

Db 58 evlglapdltsrdavmaavgtvackgrldvminnagit-mnsfsvrseedfkn-imdi 115

Qy 60 AHL-YQIDVQSEEVINGFEQIGKDVGNIDGVYHSIAFANNEDLRGFSRSGFLLAQ 118

Db 116 nvngvfngawagcmk-dakqgvilintasytgytgsiggyppqggvgigthlgre 174

Qy 119 DISS-YSLTIVAHEAKKLMPGGGSIVATTYLGGEFAVQNVNMGVAKASLEANNVKYIALD 177

Db 175 iirknirvvgavpvdtdmtkgipp-e-lledylkklpkmrkpkeeanvylflasdl 232

Qy 178 LGPNIRNAISAGPIRTLSAKGVGFNTILKEIEERAPLKRNVYDQVEVGKTAAYLLSOL 237

Db 233 asgitattisvsgavrp 249

Qy 238 SSGVTGENIHVDSGFHA 254

RESULT 11
ENTRY C24706 #type complete
TITLE nodulation protein nodG - Rhizobium meliloti plasmid
ALTERNATE_NAMES hsc protein
ORGANISM #formal_name Rhizobium meliloti
DATE 30-Jun-1988 #sequence_revision 13-Jan-1995 #text_change 05-May-1995

ACCESSIONS C24706; S07675; C24193
REFERENCE A93638
#authors DeBelle, F.; Sharma, S.B.
#journal Nucleic Acids Res. (1986) 14:7453-7472
#title Nucleotide sequence of Rhizobium meliloti RCR2011 genes involved in host specificity of nodulation.

#cross-references MUID:87016382
#accession C24706
#molecule_type DNA
#residues 1-245 #label DEB
#experimental_source strain RCR2011 symbiotic plasmid S06395
REFERENCE Fisher, R.F.; Swanson, J.A.; Mulligan, J.T.; Long, S.R.
#authors Genetics (1987) 117:191-201
#journal Extended region of nodulation genes in Rhizobium meliloti 1021. II. Nucleotide sequence, transcription start sites and protein products.

#accession S07675
#molecule_type DNA
#residues 1-245 #label FIS
#experimental_source strain 1021 symbiotic plasmid A94655

REFERENCE Horvath, B.; Kondorosi, E.; John, M.; Schmidt, J.; Toeroek, I.; Gyeorgypal, Z.; Barabas, I.; Wieneke, U.; Schell, J.; Kondorosi, A.
#authors Cell (1986) 46:335-343
#journal Organization, structure and symbiotic function of Rhizobium meliloti nodulation genes determining host specificity for alfalfa.

#cross-references MUID:86272081
#accession C24193
#status significant sequence differences
#molecule_type DNA
#cross-references GB:M14052
#experimental_source strain AK631 (a variant of strain 41)
#note the sequence reported in this reference is incorrect due to multiple frameshift errors
COMMENT This is one of several proteins that control host specificity of

root hair infection and nodulation.

GENETICS
#gene nodG: hsc
#genome plasmid
CLASSIFICATION #superfamily ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
KEYWORDS host range; nodulation
FEATURE 7-183
#domain short-chain alcohol dehydrogenase homology
#label SADH

SUMMARY #length 245 #molecular-weight 26088 #checksum 6343

Query Match 11.2%; Score 197; DB 1; Length 245;
Best Local Similarity 24.1%; Pred. No. 9.25e-10;

Matches 62; Conservative 73; Mismatches 108; Indels 14; Gaps 13;

Db 1 mfeitgrkalvtgag--aiggaiaarvihaqga-iv-glhgtqiek-letlatelgdr-v 54

Qy 1 MLNLENTYVINGIANKRSIAFGVAKVLDQAKLVFTYRKERSRKELEKLEQLNQPEA 60

Db 55 klifpanlanrdevkalgraeadilegvdlvnnagik-dglflhmadpdd-ivlevnl 112

Qy 61 HLQIDVQSEEVINGFEQIGKDVGNIDGVYHSIAFANNEDLRGFSRSGFLLAQDI 120

Db 113 tamfrltreitq-qmrrirngriinvtsvagaigppgqtnycaskagmifskslaqeia 171

Qy 121 SS-YSLTIVAHEAKKLMPGGGSIVATTYLGGEFAVQNVNMGVAKASLEANNVKYIALD 179

Db 172 trnitvncvapfi-e-sam-tdklnhkqekimvaipihrmgtgtvasavaylasdha 228

Qy 180 PDNIRNAISAGPIRTLSAKGVGFNTILKE-IEERAPLKRNVYDQVEVGKTAAYLLSOL 238

Db 229 avytgttihvngmami 245

Qy 239 SSGVTGENIHVDSGFHA 255

RESULT 12

ENTRY A42468 #type complete

TITLE 7alpha-hydroxysteroid dehydrogenase (EC 1.1.1.159) -

ORGANISM Eubacterium sp. (strain VPI 12708)

DATE 24-Jul-1992 #sequence_revision 18-Sep-1992 #text_change 23-Mar-1995

ACCESSIONS A42468; A36439

REFERENCE A42468

#authors Baron, S.F.; Franklund, C.V.; Hylemon, P.B.

#journal J. Bacteriol. (1991) 173:4558-4569

#title Cloning, sequencing, and expression of the gene coding for bile acid 7alpha-hydroxysteroid dehydrogenase from Eubacterium sp. strain VPI 12708.

#cross-references MUID:91310560

#accession A42468

#status preliminary

#molecule_type DNA

#residues 1-266 #label BAR

#cross-references GB:M58743

#note the authors translated the codon CCC for residue 95 as Phe, and CCG for residue 106 as Phe

REFERENCE A36439

#authors Franklund, C.V.; de Prada, P.; Hylemon, P.B.

#journal J. Biol. Chem. (1990) 265:9842-9849

#title Purification and characterization of a microbial, NADP-dependent bile acid 7alpha-hydroxysteroid dehydrogenase.

#cross-references MUID:90277676

#accession A36439

#status preliminary

#molecule_type protein

#residues 1-22 #label FXA

CLASSIFICATION #superfamily ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

KEYWORDS oxidoreductase

Db	238	fvnghlflvdgg	249	
QY	240	GVGTGENIHVDSG	251	
RESULT	14			
ENTRY	S34724	#type complete		
TITLE	probable oxidoreductase - Serratia marcescens			
ORGANISM	#formal_name Serratia marcescens			
DATE	13-Jan-1995	#sequence_revision 13-Jan-1995	#text_change	
ACCESSIONS	S34724			
REFERENCE	S34724			
#authors	Viejo, M.; Enfedaque, J.; Regue, M.			
#submission	submitted to the EMBL Data Library, July 1993			
#description	Bacteriocin 28b from Serratia marcescens does not present lysis nor immunity genes downstream of the structural gene.			
#accession	S34724			
##status	preliminary			
##molecule_type	DNA			
##residues	1-254	#label VIE		
##cross-references	EMBL:J25281			
SUMMARY	#length 254	#molecule-weight 26613	#checksum 4302	
Query Match	10.4%;	Score 182;	DB 9;	Length 254;
Best Local Similarity	24.3%;	Pred. No. 5,20e-08;		
Matches	63;	Conservative	70;	Mismatches 111; Indels 15; Gaps 13;
Db	1	mgdlackvfigt-ae-gqigrataerlikagcdlyfhyhdsesgpkalvalahslgqka	58	
QY	1	MLNLENTTVINGIANKRSIAFGVAKVLDQGLAKLVFTYRKERS-RKELEKLLSQLNQPE	59	
Db	59	aygya-dlidtdetircvsagaeflggidilinnvg-giv-g-rkwlgeldrafwqtvid	114	
QY	60	AHLIYQIDVDSQEVINGFPQIGKDVGNIDGVYHSIAFANMEDLGRFSETSREGFLLAQD	119	
Db	115	vmntmlnvtqalpkapngaslinlaslagraghagslyvsatkagvlwttsla	174	
QY	120	ISSYSLTIVAEHA-KKLM-PPGGSIATVTLGGSEFAQVNYVM-GVAKASLEANKVYLA	175	
Db	175	aelgehgirvnvavagli-lgttrfnhrhtqtasaeetvraipigragtpddvarafal	232	
QY	176	LDLGPNIIRVNAISAGPIFTLSAKGVGGFNTILKIEE-RA-PLKRNVDQVEVGKTAAYL	233	
Db	233	aeydgifsgatldinggi	251	
QY	234	LSDLSSGVTGENIHVDSGF	252	
RESULT	15			
ENTRY	A48950	#type complete		
TITLE	aflatoxin biosynthesis gene ver-1 - Aspergillus parasiticus			
ORGANISM	#formal_name Aspergillus parasiticus			
DATE	19-Dec-1993	#sequence_revision 19-Dec-1993	#text_change	
ACCESSIONS	A48950			
REFERENCE	A48950			
#authors	Skory, C.D.; Chang, P.K.; Cary, J.; Linz, J.E.			
#journal	Appl. Environ. Microbiol. (1992) 58:3527-3537			
#title	Isolation and characterization of a gene from Aspergillus parasiticus associated with the conversion of versicolorin A to sterigmatocystin in aflatoxin biosynthesis.			
#cross-references	MUID:93128930			
#contents	NRRL 5862			
#accession	A48950			
##status	preliminary			
##molecule_type	nucleic acid			
##residues	1-262	#label SKO		
##cross-references	NCBIN:122286; NCBIPI:122287			
##note	sequence extracted from NCBI backbone			
SUMMARY	#length 262	#molecular-weight 28368	#checksum 9318	
Query Match	10.3%;	Score 181;	DB 16;	Length 262;

WIREH

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

on: Fri Dec 5 08:13:12 1997; MasPar time 44.49 Seconds
759.327 Million cell updates/sec
Regular output not generated.

Title: >US-08-790-043A-2

Description: (1-771) from US08790043A.seq

Perfect Score: 771 1 AGCTTAAATCTTGAACACAA.....GATCCACGCAATTAATAA 771

N.A. Sequence: 771 TACAAATTAGAACTTTTGT.....CTAAGGTGCGTTAATTATT

Comp: TACAAATTAGAACTTTTGT.....CTAAGGTGCGTTAATTATT

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 82499 seqs, 21907491 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

1:backl 2:51 3:52 4:53 5:54 6:55 7:56 8:PCT90 9:PCT91
10:PCT92 11:PCT93 12:PCT94 13:PCT95 14:PCT96

Statistics: Mean 8.114; Variance 4.451; scale 1.823

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	55	7.1	7218	US-08-232	Sequence 14, Applicati	4.32e-22
2	27	3.5	215	US-08-238	Sequence 5, Applicatio	4.14e-04
3	27	3.5	242	US-08-273	Sequence 1, Applicatio	4.14e-04
4	24	3.1	215	US-08-238	Sequence 5, Applicatio	2.03e-02
5	22	2.9	2894	US-08-483	Sequence 1, Applicatio	2.43e-01
6	22	2.9	2894	US-08-472	Sequence 1, Applicatio	2.43e-01
7	22	2.9	2894	US-08-278	Sequence 1, Applicatio	2.43e-01
8	21	2.7	74	PCT-US95-1	Sequence 100, Applicat	8.12e-01
9	21	2.7	74	PCT-US95-1	Sequence 98, Applicati	8.12e-01
10	21	2.7	81	PCT-US95-1	Sequence 92, Applicati	8.12e-01
11	21	2.7	81	PCT-US95-1	Sequence 1, Applicatio	8.12e-01
12	21	2.7	774	PCT-US91-0	Sequence 1, Applicatio	8.12e-01
13	21	2.7	824	US-08-158	Sequence 1, Applicatio	8.12e-01
14	21	2.7	75	PCT-US95-1	Sequence 99, Applicati	2.63e+00
15	20	2.6	81	PCT-US95-1	Sequence 98, Applicati	2.63e+00
16	20	2.6	82	PCT-US95-1	Sequence 97, Applicati	2.63e+00
17	20	2.6	579	PCT-US96-0	Sequence 1664, Applic	2.63e+00
18	20	2.6	1265	US-08-182	Sequence 5, Applicatio	2.63e+00
19	20	2.6				

20	2.6	1323	7	US-08-307-	Sequence 36, Applicati	2.63e+00
21	20	1990	1	5518916-5	Patent No. 5518916.	2.63e+00
22	20	1990	1	5171685-5	Patent No. 5171685.	2.63e+00
23	20	1990	1	5518916-5	Patent No. 5518916.	2.63e+00
24	20	4383	1	5177307-1	Patent No. 5177307.	2.63e+00
25	20	4383	1	5175095-4	Patent No. 5175095.	2.63e+00
26	20	8082	11	PCT-US93-0	Sequence 28, Applicati	2.63e+00
27	20	14176	7	US-08-307-	Sequence 14, Applicati	2.63e+00
28	20	14176	7	US-08-307-	Sequence 1, Applicatio	2.63e+00
29	19	65	7	US-08-471-	Sequence 145, Applicat	8.26e+00
30	19	66	13	PCT-US95-1	Sequence 93, Applicati	8.26e+00
31	19	58	7	US-08-471-	Sequence 143, Applicati	8.26e+00
32	19	74	13	PCT-US95-1	Sequence 94, Applicati	8.26e+00
33	19	74	13	PCT-US95-1	Sequence 100, Applicat	8.26e+00
34	19	2.5	448	US-08-702-	Sequence 7, Applicatio	8.26e+00
35	19	2.5	774	PCT-US91-0	Sequence 1, Applicatio	8.26e+00
36	19	2.5	1079	US-07-809-	Sequence 1, Applicatio	8.26e+00
37	19	2.5	1079	US-07-809-	Sequence 1, Applicatio	8.26e+00
38	19	2.5	1561	PCT-US92-0	Sequence 25, Applicati	8.26e+00
39	19	2.5	2336	PCT-US92-0	Sequence 1, Applicatio	8.26e+00
40	19	2.5	2339	PCT-US92-0	Sequence 2, Applicatio	8.26e+00
41	19	2.5	2504	US-08-484-	Sequence 15, Applicati	8.26e+00
42	19	2.5	2712	PCT-US95-0	Sequence 37, Applicati	8.26e+00
43	19	2.5	3104	PCT-US95-0	Sequence 66, Applicati	8.26e+00
44	19	2.5	4608	US-07-718-	Sequence 11, Applicati	8.26e+00
45	19	2.5	14311	PCT-US96-0	Sequence 7, Applicatio	8.26e+00

ALIGNMENTS

RESULT 1
ID US-08-232-463-14 STANDARD; DNA; UNC; 7218 BP.
AC xxxxxx
DT 01-JAN-1900
DE Sequence 14, Application US/08232463.
Sequence 14, Application US/08232463
CC Patent No. 5670367
CC GENERAL INFORMATION:
CC APPLICANT: DORNER, F.
CC APPLICANT: SCHEIFLINGER, F.
CC APPLICANT: FALKNER, F. G.
CC TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
CC NUMBER OF SEQUENCES: 52
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Foley & Lardner
CC STREET: 1800 Diagonal Road, Suite 500
CC CITY: Alexandria
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22313-0299
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/232.463
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC PRIOR APPLICATION NUMBER: US/07/935.313
CC FILING DATE:
CC APPLICATION NUMBER: EP 91 114 300.6
CC FILING DATE: 26-AUG-1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: BENT, Stephen A.
CC REGISTRATION NUMBER: 29,768
CC REFERENCE/DOCKET NUMBER: 30472/114 IMMU
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703)836-9300
CC TELEFAX: (703)683-4109
CC TELEX: 899149
CC INFORMATION FOR SEQ ID NO: 14:

```

CC CLASSIFICATION: 800
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Bastian, Kevin L.
CC REGISTRATION NUMBER: 34,774
CC REFERENCE/DOCKET NUMBER: 2307E-540
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (415) 543-9600
CC TELEFAX: (415) 543-5043
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 215 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC FEATURE:
CC NAME/KEY: misc.feature
CC LOCATION: 1..215
CC OTHER INFORMATION: /standard_name= "Deduced amino acid
CC OTHER INFORMATION: sequence of PGIP from bean."
CC SQ Sequence 215 BP; 15 A; 8 C; 25 G; 26 T; 141 other;

Query Match 3.5%; Score 27; DB 6; Length 215;
Best Local Similarity 17.8%; Pred. No. 4.14e-04;
Matches 26; Conservative 55; Mismatches 63; Indels 2; Gaps 2;

Db 43 VCOTDITRYVNDSDGHKYSANVYGGNNVGAAKTHYYTHNVSGADSKTVDTSYNASG 102
QY 368 ACCTACATGAACATTTAGTGGCTCATGAAGCTAAAAAATTAATGCCAAGGTTGGTAGCATTG 427
Db 103 TSSSSNGGTGDN-RSG-ADSYGSSKTMATSRNRTGKTANNAVDSRNMGDASVSDKNYKKH 160
QY 428 TTGCAACAACATATTTAGTGGCGAATTCGCATTCAAAATTAATATGTGATGGGTGTTG 487
Db 161 AKNSADGKVGSKNGDNRNRYGTGK 186
QY 488 CTAAGCGAGCTTAGAAGCAATGTT 513

RESULT 3
ID US-08-273-846-1 STANDARD; RNA; UNC; 290 BP.
AC XXXXX
DT 01-JAN-1900
DE Sequence 1, Application US/08273846.
CC Sequence 1, Application US/08273846
CC Patent No. 5641863
CC GENERAL INFORMATION:
CC APPLICANT: SCHREIBER, ALAN D.
CC APPLICANT: PARK, JONG-GU
CC TITLE OF INVENTION: METHODS OF STIMULATING PHAGOCYTOSIS
CC NUMBER OF SEQUENCES: 7
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: NIXON & VANDERHYE P.C.
CC STREET: 1100 NORTH GLEBE ROAD
CC CITY: ARLINGTON
CC STATE: VIRGINIA
CC COUNTRY: U.S.A.
CC ZIP: 22201-4714
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/273,846
CC FILING DATE: 12-JUL-1994
CC CLASSIFICATION: 530
CC ATTORNEY/AGENT INFORMATION:
CC NAME: WILSON, MARY J.
CC REGISTRATION NUMBER: 32,955
CC REFERENCE/DOCKET NUMBER: 555-29
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703) 816-4000

```


Cp 450 GCCACCTAAATATGTTGTCACAAATGCTACACCTTCTGGCAATTAAT 402

RESULT 5

ID US-08-483-859-1 STANDARD; DNA; UNC; 2894 BP.

AC xxxxxx

DT 01-JAN-1900

DE Sequence 1, Application US/08483859.

CC Sequence 1, Application US/08483859.

CC Patent No. 5656436

CC GENERAL INFORMATION:

CC APPLICANT: LOOSMORE, Sheena M.

CC APPLICANT: YANG, Yan-Ping

CC APPLICANT: CHONG, Pele

CC APPLICANT: COMEN, Raymond P.

CC APPLICANT: KLEIN, Michel H.

CC TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with

CC TITLE OF INVENTION: Reduced Protease Activity

CC NUMBER OF SEQUENCES: 23

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Sim & McBurney

CC STREET: Suite 701, 330 University Avenue

CC CITY: Toronto

CC STATE: Ontario

CC COUNTRY: Canada

CC ZIP: M5G 1R7

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: PatentIn Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/483,859

CC FILING DATE: 07-JUN-1995

CC CLASSIFICATION: 435

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/296,149

CC FILING DATE: 26-AUG-1994

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/278,091

CC FILING DATE: 21-JUL-1994

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Stewart, Michael I.

CC REGISTRATION NUMBER: 24,973

CC REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (416) 595-1155

CC TELEFAX: (416) 595-1163

CC INFORMATION FOR SEQ ID NO: 1:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 2894 base pairs

CC TYPE: nucleic acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC Sequence 2894 BP; 897 A; 570 C; 515 G; 912 T; 0 other;

Query Match 2.9%; Score 22; DB 7; Length 2894;

Best Local Similarity 75.0%; Pred. No. 2.43e-01;

Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 2675 CGCTTGTGTTGCGTCAATCTTCTCAATGATGTTGTCACGCCCTT 2718

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Cp 180 CGCTTGTGTTGATTAATGTTCTCAATAATTTTCAAGCTCTT 137

RESULT 6

ID US-08-472-173-1 STANDARD; DNA; UNC; 2894 BP.

AC xxxxxx

DT 01-JAN-1900

DE Sequence 1, Application US/08472173.

CC Sequence 1, Application US/08472173.

CC Patent No. 5665353

CC GENERAL INFORMATION:

CC APPLICANT: LOOSMORE, Sheena M

CC APPLICANT: YANG, Yan-Ping

CC APPLICANT: CHONG, Pele

CC APPLICANT: COMEN, Raymond P.

CC APPLICANT: KLEIN, Michel H.

CC TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with

CC TITLE OF INVENTION: Reduced Protease Activity

CC NUMBER OF SEQUENCES: 23

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Sim & McBurney

CC STREET: Suite 701, 330 University Avenue

CC CITY: Toronto

CC STATE: Ontario

CC COUNTRY: Canada

CC ZIP: M5G 1R7

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: PatentIn Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/472,173

CC FILING DATE: 07-JUN-1995

CC CLASSIFICATION: 435

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/296,149

CC FILING DATE: 26-AUG-1994

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: US 08/278,091

CC FILING DATE: 21-JUL-1994

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Stewart, Michael I.

CC REGISTRATION NUMBER: 24,973

CC REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (416) 595-1155

CC TELEFAX: (416) 595-1163

CC INFORMATION FOR SEQ ID NO: 1:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 2894 base pairs

CC TYPE: nucleic acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC Sequence 2894 BP; 897 A; 570 C; 515 G; 912 T; 0 other;

Query Match 2.9%; Score 22; DB 7; Length 2894;

Best Local Similarity 75.0%; Pred. No. 2.43e-01;

Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 2675 CGCTTGTGTTGCGTCAATCTTCTCAATGATGTTGTCACGCCCTT 2718

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Cp 180 CGCTTGTGTTGATTAATGTTCTCAATAATTTTCAAGCTCTT 137

RESULT 7

ID US-08-278-091-1 STANDARD; DNA; UNC; 2894 BP.

AC xxxxxx

DT 01-JAN-1900

DE Sequence 1, Application US/08278091.

CC Sequence 1, Application US/08278091.

CC Patent No. 5506139

CC GENERAL INFORMATION:

CC APPLICANT: LOOSMORE, Sheena M

CC APPLICANT: YANG, Yan-Ping

CC APPLICANT: CHONG, Pele

CC APPLICANT: COMEN, Raymond P.

CC APPLICANT: KLEIN, Michel H.

CC TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with

CC TITLE OF INVENTION: Reduced Protease Activity

CC NUMBER OF SEQUENCES: 23

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Sim & McBurney

CC STREET: Suite 701, 330 University Avenue

[illegible]

W E R E H
***** (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
on: Fri Dec 5 07:57:33 1997; MasPar time 100.32 Seconds
802.071 Million cell updates/sec

Title: >US-08-790-043A-2
Description: (1-771) from US08790043A.seq
Perfect Score: 771
N.A. Sequence: 1 ATGTTAAATCTTGAACAA.....GATTCACGCAATTAATAA 771
Comp: TACAATTTAGAACTTTGTT.....CTAAGTGCGTTAATTAT

Scoring table: TABLE default
Gap 6
Nmatch STD : Dbase 0; Query 0
Searched: 142080 seqs, 52183452 bases x 2
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: n-genseq28
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29
Statistics: Mean 8.559; Variance 5.316; scale 1.610
pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES				Pred. No.	
Result No.	Score	Query Match	Description	Pred. No.	
1	63	8.2	1047 2 Q10572 Human Natriuretic Pep	3.58e-23	
2	52	6.7	1047 2 Q10572 Human Natriuretic Pep	1.56e-16	
3	40	5.2	91 9 Q51746 Oligonucleotide probe	1.42e-09	
4	38	4.9	204 1 N81164 Base substituted E.co	1.86e-08	
5	37	4.8	91 9 Q51746 Oligonucleotide probe	6.67e-08	
6	34	4.4	204 1 N81164 Base substituted E.co	2.90e-06	
7	33	4.3	114 12 Q70467 Generic DNA sequence	9.99e-06	
8	32	4.2	114 12 Q70465 Generic DNA sequence	3.41e-05	
9	31	4.0	114 12 Q70468 Generic DNA sequence	1.15e-04	
10	30	3.9	114 12 Q70468 Generic DNA sequence	3.81e-04	
11	30	3.9	114 12 Q70466 Generic DNA sequence	3.81e-04	
12	28	3.6	114 12 Q70470 Generic DNA sequence	4.05e-03	
13	25	3.2	114 12 Q70465 Generic DNA sequence	1.25e-01	
14	24	3.1	114 12 Q70470 Generic DNA sequence	3.76e-01	
15	24	3.1	114 12 Q70469 Generic DNA sequence	3.76e-01	

C	16	24	3.1	114 12	Q70472	Generic DNA sequence	3.76e-01
C	17	23	3.0	36 2	Q11195	Ballast Constituent c	1.11e+00
C	18	23	3.0	39 7	Q51787	Mixed oligonucleotide	1.11e+00
C	19	23	3.0	114 12	Q70468	Generic DNA sequence	1.11e+00
C	20	22	2.9	30 7	Q48602	Mixed oligonucleotide	3.23e+00
C	21	22	2.9	33 7	Q48603	Mixed oligonucleotide	3.23e+00
C	22	22	2.9	33 2	Q11197	Ballast Constituent c	3.23e+00
C	23	22	2.9	114 12	Q70467	Generic DNA sequence	3.23e+00
C	24	22	2.9	114 12	Q70466	Generic DNA sequence	3.23e+00
C	25	22	2.9	114 12	Q70473	Generic DNA sequence	3.23e+00
C	26	22	2.9	498 3	N50034	Sequence encoding new	3.23e+00
C	27	22	2.9	501 3	N50027	Sequence encoding new	3.23e+00
C	28	22	2.9	501 3	N50024	Sequence encoding new	3.23e+00
C	29	22	2.9	501 3	N50029	Sequence encoding new	3.23e+00
C	30	22	2.9	501 3	N50030	Sequence encoding new	3.23e+00
C	31	22	2.9	501 3	N50033	Sequence encoding new	3.23e+00
C	32	22	2.9	501 3	N50032	Sequence encoding new	3.23e+00
C	33	22	2.9	501 3	N50023	Sequence encoding new	3.23e+00
C	34	22	2.9	501 3	N50028	Sequence encoding new	3.23e+00
C	35	22	2.9	501 3	N50031	Sequence encoding new	3.23e+00
C	36	22	2.9	2894 18	T10423	H. influenzae SB33 hi	3.23e+00
C	37	22	2.9	3871 2	N71302	HSV-1 gB and surround	3.23e+00
C	38	21	2.7	74 21	T13613	DC43 TSAR library gen	9.13e+00
C	39	21	2.7	81 21	T13611	DC43 TSAR library gen	9.13e+00
C	40	21	2.7	453 8	Q49218	Reverse translated Ap	9.13e+00
C	41	21	2.7	501 3	N50025	Sequence encoding new	9.13e+00
C	42	21	2.7	534 12	Q73883	Borrelia 2591 antigen	9.13e+00
C	43	21	2.7	774 3	Q22981	Gelonin toxin DNA	9.13e+00
C	44	21	2.7	2403 21	T12562	S. aureus topoisomeras	9.13e+00
C	45	21	2.7	5746 24	T09225	Partial sequence of v	9.13e+00

ALIGNMENTS

RESULT 1
ID Q10572 standard; DNA; 1047 BP.
AC Q10572;
DT 09-APR-1991 (first entry)
DE Human Natriuretic Peptide Receptor B.
KW NPB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
hyperaldosteronism; glaucoma; guanyl cyclase.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= signal sequence
FT Protein 12
FT /label= mature NPB
FT Domain 23..455
FT /label= extracellular domain
FT /note= "binds natriuretic peptides A,B and C]"
FT Domain 456..456
FT /label= transmembrane domain
FT Domain 479..1047
FT /label= cytoplasmic domain
FT /note= "GC and protein kinase activity"
FT Modified-site 24..26
FT /label= N-glycos_site
FT Modified-site 35..37
FT /label= N-glycos_site
FT Modified-site 161..163
FT /label= N-glycos_site
FT Modified-site 195..197
FT /label= N-glycos_site
FT Modified-site 244..246
FT /label= N-glycos_site
FT Modified-site 277..279
FT /label= N-glycos_site
FT Modified-site 349..351
FT /label= N-glycos_site
FT Modified-site 600..602
FT /label= N-glycos_site
PN W09100292-A.
PD 10-JAN-1991.

22-JUN-1990; U03586.
23-JUN-1989; US-370673.
PA (GETH) GENENTECH INC.
PI Chang M, Goeddel D, Lowe D;
WPI; 91-036711/05.
DR N-PSDB; 010324.
PT Natriuretic protein receptor B - for diagnosis and treatment of
PT kidney failure, heart failure, hyperaldosteronism, glaucoma etc.
PS Claim 3; Fig 1; 49pp; English.
CC The sequence was derived from the DNA encoding natriuretic peptide
CC receptor B, NPRB, having guanyl cyclase (GC) activity and protein
CC kinase activity. The DNA can be inserted into expression vectors
CC for the prodn. of the protein, opt. after being mutated to produce
CC NPRB analogues. The protein has a mol wt. of 115 kD (calculated Mr=
CC 114,952). The protein (or variants) can be used in treatment of
CC natriuretic peptide disorders, and also to isolate peptides using
CC affinity chromatography. Antibodies with affinity for NPRB can
CC also be prepd.
SQ Sequence 1047 BP; 87 A; 15 C; 83 G; 51 T;

Try Match 8.2%; Score 63; DB 2; Length 1047;
St Local Similarity 11.4%; Pred. No. 3.58e-23;
Matches 65; Conservative 155; Mismatches 341; Indels 7; Gaps 7;

D b 78 ynannsadvknknyhdnnngncvynnaasvarnashwnrnttgagvasgsakndhyr 137
Q y : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
168 TCACCCAGAACGCCGCTATAT -CAAAATTGATTTCAAACCGCATGAAGAGGTATAATG 226
D b 138 tnvrtgnsankngnnvtnhgghnwitar-aannyndartddrhnytntngvnnaangsn 196
Q y : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
227 GTTTTGAGCAAAATGGTFAAAGATTGGCAATATTGATGGTGATATCATTCATCGCAT 286
D b 197 svnhvyarnngnnnathnrangrnrvncgnnmhmnnnnanrrntngdyvynynd 256
Q y TTGCTAATATGGAAGACTTAGCCGGACGCTTTTCTGAAACTTCAGTGGAAGCTTCTTGT 346
D b 257 vngnsnragnratgrtnwdrtrnnanaurnantvntvyrnnnnnnnnnnnnr 316
Q y : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
347 TAGCTCAAGACATTAGTTCTTACTCATTACAAATTGGCTCATGAAGCTAAAAAATAA 406
D b 317 arndngvngnsmnnnagcnydgnnnyanvnnntnnnggttrndgrnvrnkmgryrhgv 376
Q y 407 TGCACGAAGGTGTAGCATTTGTGCACAACACATATTAGGTGGCAATTCGCAGTTCAA 466
D b 377 tgnvmdkndrdtndvnamgdndsgdnnaahysganknnwrgnrvnnvkgnnsdn 436
Q y : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
467 ATTATAATGTGA-TGGGTGTGCTAAGCGAGCTTAGAGAACAATG-TTAAATATTAGC 524
D b 437 nncandnddsckdtnnstnanvangtgnntmngvssnnnrkmmnkknasmwrrw 496
Q y : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
525 ATTAGACTTAGGTCCTGATAATATTCGCGTTAATGCAATTTTCAG-CTGGTC-CATTCGT 582
D b 497 nnnngnsnryhkagsrntnsrgssygsmtahg-kynnnaentghknvnavkhvknk 555
Q y 583 ACATTAAAGTCAAAAGGTGGGTGGTGTTCATACAAATTCCTTTAAAGAAATCGACAGCGT 642
D b 556 krnantrnnnnkhnrdvnnhntrngacndnnncvtncyrsgndnnndnnndnnn 615
Q y 643 GCACCTTTAAACGTAACGTTGATCAAGTAGAAGTAGGTAAACACAGCGGCTTACTTRTA 702
D b 616 dwmnrysnndnvkgmannhnsnsshg 643
Q y : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
703 AGTGACTTATCAAGTGGCGGCTTACAGGTG 730

RESULT 2
ID Q10572 standard; DNA; 1047 BP.
AC Q10572:
DE DT 09-APR-1991 (first entry)
DT Human Natriuretic Peptide Receptor B.
KW NPRB; ANP; BNP; CNP; kidney failure; heart failure; protein kinase;
KW hyperaldosteronism; glaucoma; guanyl cyclase.
OS Homo sapiens.

RESULT	4
ID	N81164 standard; DNA; 204 BP.
AC	N81164;
DT	08-NOV-1990 (first entry)
DE	Base substituted E.coli beta-galactosidase alpha-fragment.
KW	E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS	Escherichia coli.
FH	Key
FT	Location/Qualifiers
FT	misc_feature 19..69
FT	/*tag= a
FT	/function-multiple cloning site

comprising a binding domain and an effector domain
 PS Disclosure; Page 35; 255pp; English.
 CC Q70465 is a generic DNA sequence used to generate random TSAR (Totally
 CC Synthetic Affinity Reagents) peptides. This generic formula can also be
 CC represented as follows: X(NNB)6(TGC)(NNB)14(TGC)(NNB)3Y. X
 CC and Y are flanking restriction sites (X is not the same as Y) that are
 CC not specified further. Other generic sequences are shown in Q70466-68.
 CC Other specific peptides generated by these generic sequences are shown in
 CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
 CC comprising at least two functional regions - a binding domain with
 CC affinity for a ligand and a second effector peptide portion that is
 CC chemically or biologically active. They may further comprise a linker
 CC peptide between the 2 domains. The oligonucleotides are also designed so
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned
 CC in, or flanking, the unpredicted or variant residues. These residues
 CC confer some degree of conformational rigidity to the peptides. The TSARs
 CC or compns. comprising a TSAR binding domain can be used in vivo to
 CC deliver a chemically or biologically active moiety, eg. metal ion,
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
 CC cell. They can also replace the function of macromolecules, eg.
 CC monoclonal or polyclonal antibodies and therefore circumvent the need
 CC for complex methods of hybridoma formation or in vivo antibody
 CC production. The TSARs are easily characterised and have designed
 CC activity allowing direct and rapid detection in a screening process.
 CC Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;
 SQ

Query Match 4.2%; Score 32; DB 12; Length 114;
 Best Local Similarity 2.7%; Pred. NO. 3.4le-05;
 Matches 3; Conservative 33; Mismatches 76; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbgtcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnn 62
 Cp 667 GATCAACGGTTACGTTAAAGGTGCGACGCTCTCGATTCTTAAAGATTGTTAAGAAC 608

Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
 Cp 607 CACCACACCTTTTGACCTTAATGTACGGATTGGACGACGCTGAATGTCATT 556

RESULT 9
 ID Q70469 standard; DNA; 114 BP.
 AC Q70469;
 DT 07-APR-1995 (first entry)
 DE Generic DNA sequence to generate a random TSAR peptide library.
 KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 PH Key Location/Qualifiers
 FT misc_feature 55..60
 FT /*tag= a
 FT /note= "this sequence represents 'Z'; Z can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 PN WO9418318-A.
 PD 18-AUG-1994.
 PF 01-FEB-1994; U00977.
 PR 01-FEB-1994; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PA (UYNC-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR WPI: 94-279739/34.
 DR P-PSDB; R65154.
 PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 PS Disclosure; Page 35; 255pp; English.
 CC Q70469 is a generic DNA sequence used to generate random TSAR peptide
 CC This generic formula can be represented as follows: X(TGC)(NNB)10-
 CC (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TGC)Y. X and Y are flanking restriction
 CC sites (X is not the same as Y) that are not specified further. Other
 CC sequence generates peptides that are cloverleaf in structure. Other
 CC generic sequences are shown in Q70465-68. Other specific peptides

generated by these generic sequences are shown in R65150-54. TSARs are
 concatenated heterofunctional proteins or peptides, comprising at least
 two functional regions - a binding domain with affinity for a ligand and
 a second effector peptide portion that is chemically or biologically
 active. They may further comprise a linker peptide between the 2 domains.
 The oligonucleotides are also designed so that the expressed peptide
 contains 2 or 4 cysteine residues positioned in, or flanking, the
 unpredicted or variant residues. These residues confer some degree of
 conformational rigidity to the peptides. The TSARs or compns. comprising
 a TSAR binding domain can be used in vivo to deliver a chemically or
 biologically active moiety, eg. metal ion, radioisotope, peptide, toxin
 or enzyme, to the specific target or on the cell. They can also replace
 the function of macromolecules, eg. monoclonal or polyclonal antibodies
 and therefore circumvent the need for complex methods of hybridoma
 formation or in vivo antibody production. The TSARs are easily
 characterised and have designed activity allowing direct and rapid
 detection in a screening process.
 Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;
 SQ

Query Match 4.0%; Score 31; DB 12; Length 114;
 Best Local Similarity 5.4%; Pred. NO. 1.15e-04;
 Matches 6; Conservative 30; Mismatches 75; Indels 0; Gaps 0;

Db 3 cnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 62
 Cp 676 CTTCTACTTGATCAACGTTACGTTTAAAGGTGCGACGCTCTTCGATTCTTTAAGATTG 617

Db 63 bnnbgtcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 113
 Cp 616 TATTGAACACCCACACACCTTTTGACCTTAATGTACGGATTGGACGACGCTG 566

RESULT 10
 ID Q70468 standard; DNA; 114 BP.
 AC Q70468;
 DT 05-APR-1995 (first entry)
 DE Generic DNA sequence to generate a random TSAR peptide library.
 KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
 KW effector domain; concatenated heterofunctional protein; linker;
 KW direct; rapid; detection; screening; treatment; generic; ss.
 OS Synthetic.
 PH Key Location/Qualifiers
 FT misc_feature 55..60
 FT /*tag= a
 FT /note= "this sequence represents 'Z'; Z can be a
 FT sequence of 6, 9 or 12 nucleotides (see
 FT comments)"
 PN WO9418318-A.
 PD 18-AUG-1994.
 PF 01-FEB-1994; U00977.
 PR 01-FEB-1994; US-013416.
 PR 30-DEC-1993; US-176500.
 PR 31-JAN-1994; US-189331.
 PA (UYNC-) UNIV NORTH CAROLINA.
 PI Fowlkes DM, Kay BK;
 DR WPI: 94-279739/34.
 DR P-PSDB; R65154.
 PT Identifying proteins or peptide(s) which bind a ligand - by
 PT screening a recombinant vector library expressing fusion proteins
 PT comprising a binding domain and an effector domain
 PS Disclosure; Page 35; 255pp; English.
 CC Q70468 is a generic DNA sequence used to generate random TSAR (Totally
 CC Synthetic Affinity Reagents) peptides. This generic formula can also be
 CC represented as follows: X(NNB)11(TGC)(NNB)6Z(NNB)7(TGC)(NNB)10Y. X
 CC and Y are flanking restriction sites (X is not the same as Y) that are
 CC not specified further. Other generic sequences are shown in Q70466-68.
 CC Other specific peptides generated by these generic sequences are shown in
 CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
 CC comprising at least two functional regions - a binding domain with
 CC affinity for a ligand and a second effector peptide portion that is
 CC chemically or biologically active. They may further comprise a linker
 CC peptide between the 2 domains. The oligonucleotides are also designed so
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned

[illegible]

100

Search completed: Fri Dec 5 07:59:19 1997
Job time : 106 secs.

WAPREH (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligent, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

on: Fri Dec 5 07:44:28 1997; MasPar time 750.09 Seconds
1128.898 Million cell updates/sec
Similar output not generated.

Title: >US-08-790-043A-2
Description: (1-771) from US08790043A.seq
Perfect Score: 771
N.A. Sequence: 1 ATGTTAAATCTTGAACAA.....GATTCACGCAATTAATAA 771
Comp: TACAATTAGAACTTTGTT.....CTAAGTGCGTTAATTAT

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 362067 seqs, 549138275 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: emb1-new3
1: BCT 2: FUN 3: GEN1 4: GEN2 5: HTG1 6: HTG2 7: HUM 8: INV
9: ORG 10: MAM 11: VRT 12: PLN 13: PRO 14: ROD 15: SYN 16: UNC
17: VIR

Database: genbank99
18: BCT1 19: BCT2 20: BCT3 21: BCT4 22: BCT5 23: BCT6 24: BCT7
25: BCT8 26: BCT9 27: BCT10 28: BCT11 29: GEN1 30: GEN2
31: GEN3 32: HTG1 33: HTG2 34: HTG3 35: INV1 36: INV2 37: INV3
38: INV4 39: INV5 40: INV6 41: INV7 42: INV8 43: INV9 44: INV10
45: INV11 46: MAM1 47: MAM2 48: MAM3 49: VRT1 50: VRT2 51: VRT3
52: VRT4 53: PAT1 54: PAT2 55: PAT3 56: PAT4 57: PAT5 58: PHG
59: PLN1 60: PLN2 61: PLN3 62: PLN4 63: PLN5 64: PLN6 65: PLN7
66: PLN8 67: PLN9 68: PLN10 69: PLN11 70: PRI1 71: PRI2
72: PRI3 73: PRI4 74: PRI5 75: PRI6 76: PRI7 77: PRI8 78: PRI9
79: PRI10 80: PRI11 81: PRI12 82: PRI13 83: PRI14 84: PRI15
85: ROD1 86: ROD2 87: ROD3 88: ROD4 89: ROD5 90: ROD6 91: ROD7
92: ROD8 93: SYN 94: SYN 95: UNA 96: VRL1 97: VRL2 98: VRL3
99: VRL4 100: VRL5 101: VRL6 102: VRL7 103: VRL8 104: VRL9
105: VRL10

Database: genbank-new3
106: BCT 107: GEN1 108: GEN2 109: HTG1 110: HTG2 111: INV
112: MAM 113: VRT 114: PHG 115: PLN 116: PRI1 117: PRI2
118: ROD 119: SYN 120: UNA 121: VRL

Database: u-emb150_99
122: part1

Statistics: Mean 10.783; Variance 5.167; scale 2.087

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	110	14.3	4142	18	ANABIFA	Anabaena sp. sequence	3.86e-64
2	89	11.5	122056	20	D90902	Synechocystis sp. PCC	4.50e-47
3	66	8.6	2123	25	PM038482	Proteus mirabilis glu	5.34e-29
C 4	56	7.3	10111	23	HI032846	Haemophilus influenza	1.87e-21
C 5	56	7.3	10111	13	HI032846	Haemophilus influenza	1.87e-21
C 6	43	5.6	1366	21	ECENVM	Escherichia coli shor	4.14e-12
7	43	5.6	1366	21	ECENVM	Escherichia coli shor	4.14e-12
C 8	43	5.6	10477	21	ECAE000227	Escherichia coli from	4.14e-12
C 9	43	5.6	17937	106	D90766	E.coli genomic DNA, K	4.14e-12
C 10	43	5.6	17937	20	D90766	E.coli genomic DNA, K	4.14e-12
C 11	43	5.6	17769	20	D90767	S.typhimurium envm pr	4.69e-10
C 12	40	5.2	1656	27	STVENVM	Oxytricha fallax 57kD	1.26e-03
13	30	3.9	354	8	OFU09259	Oxytricha fallax 57kD	1.26e-03
14	30	3.9	354	111	OFU09259	Oxytricha fallax 57kD	1.26e-03
15	30	3.9	1405	25	MYCSEQB	Mycoplasma bovine gro	1.26e-03
16	28	3.6	1456	25	MYCSEQE	Mycoplasma mycoides g	1.92e-02
17	28	3.6	1463	25	MYCSEQF	Mycoplasma mycoides g	1.92e-02
18	28	3.6	1524	25	MYCSEQA	Mycoplasma mycoides g	1.92e-02
19	27	3.5	215	57	I28278	Sequence 5 from paten	7.21e-02
20	27	3.5	2652	18	ACGLUDEH	Acinetobacter calcoac	7.21e-02
21	27	3.5	2652	18	ACCGD	A. calcoaceticus quino	7.21e-02
22	26	3.4	1071	97	HAU07260	Helicoverpa armigera	2.64e-01
23	25	3.2	584	39	DDAHCH	D. discoideum mRNA fr	9.38e-01
24	25	3.2	1199	65	S74167	Wh1-switch-regulated	9.38e-01
25	25	3.2	1315	61	DDIAHHA	Slime mold (D.discoso	9.38e-01
26	25	3.2	1349	67	SCYNL203C	S.cerevisiae chromoso	9.38e-01
C 27	25	3.2	1881	69	YSCSPX19A	S.cerevisiae SPX19 an	9.38e-01
28	25	3.2	3124	18	BBU28760	Borrelia burgdorferi	9.38e-01
29	25	3.2	4310	75	HSLAZ3	H.sapiens LAZ3/BCL6 g	9.38e-01
30	25	3.2	35161	66	SCXIVL	S.cerevisiae DNA of c	9.38e-01
31	24	3.1	215	57	I28278	Sequence 5 from paten	3.23e+00
32	24	3.1	1182	79	HUM33DPTP	Human 33-kDa phototra	3.23e+00
33	24	3.1	1485	25	MYCSEQC	Mycoplasma capricolum	3.23e+00
34	24	3.1	26969	38	CET07F10	Caenorhabditis elegans	3.23e+00
35	23	3.0	1983	25	MYC1623SR	Mycoplasma sp. 16S ri	1.07e+01
36	23	3.0	3661	26	SDFNBBA	S.dysgalactiae fnbB g	1.07e+01
37	23	3.0	3684	26	SEFNBP	S.equisimilis (Sel65)	1.07e+01
38	23	3.0	6969	44	SCMVH	Schistosoma mansoni m	1.07e+01
39	23	3.0	25284	38	CEW01A8	Caenorhabditis elegans	1.07e+01
C 40	23	3.0	34944	111	TG087145	Toxoplasma gondii chl	1.07e+01
C 41	23	3.0	34944	9	CHTGGU871	Toxoplasma gondii chl	1.07e+01
42	23	3.0	130690	20	D90901	Synechocystis sp. PCC	1.07e+01
43	22	2.9	4800	19	CLOAAD	Clostridium acetobuty	3.44e+01
44	22	2.9	5040	18	BARPLRPOG	B. aphidicola rpl, rp	3.44e+01
C 45	22	2.9	36258	36	CEF29F11	Caenorhabditis elegans	3.44e+01

ALIGNMENTS

RESULT 1 ANABIFA 4142 bp DNA BCT 18-AUG-1993
LOCUS Anabaena sp. sequence-specific DNA binding protein (bifa) gene,
complete cds.
DEFINITION
ACCESSION L10036
NID g142008
KEYWORDS DNA binding protein; bifa gene.
SOURCE Anabaena sp. (library: PCC 7120) DNA.
ORGANISM Anabaena sp.
REFERENCE 1 (bases 1 to 4142)
AUTHORS Wei, T.-F., Ramasubramanian, T.S., Pu, F. and Golden, J.W.
TITLE Anabaena sp. strain PCC 7120 bifa gene encoding a sequence-specific
DNA-binding protein cloned by in vivo transcriptional interference
selection
J. Bacteriol. 175, 4025-4035 (1993)
MEDLINE 93308081
FEATURES Location/Qualifiers
1..4142
/organism="Anabaena sp."
/tissue_lib="PCC 7120"

CDS

```

complement(<1..498)
/note="No homologous sequence was found in GenBank"
/codon_start=1
/label=ORF5
/db_xref="PID:g289134"
/transl_table=11
/translation="MTTGVLIAAILLGGVATVGVDRIGTRVGRKARLSLFLNRPKNT
AVLVTLLGGVLSATLAILFTADEGLRKGVELEDIQDLRQKREQLKVAEOKTQV
EIERKNVQNELETTTRDKKQVETORDQAKKEKLKAQDLAQTAQYQRTQSLRGQVVT
QYQKAI"
complement(828..1499)
/label="bifa"
/label="bifa"
/codon_start=1
/product="DNA-binding protein"
/db_xref="PID:g142009"
/transl_table=11
/translation="MIVTQKALANVFROMATGAPPPVETFERNKTIFFPPDPAERV
YFLKLGAVLKRVEAGEITVALLRENSVGLSLLTGNKSDRFYHAYATPVVELLS
APIQVEQALKENPELSMLRLGLSSRIQTEMETLHAHDMGSRVLSLILCRDF
GVPCADGTTIDLKLSHQIAEAGISRTVTVTRLLGDUREKKMISIHKKKITVHKPVTL
SRQT"
2110..2904
/note="Shows 70.2% similarity and 48.6% identity to the
EnvM protein of Salmonella typhimurium"
/codon_start=1
/label=ORF3
/db_xref="PID:g142010"
/transl_table=11
/translation="MTTKISMLNTGKNALVTGIANNRSIAWGIAQOLHAAGANLGIT
YLPDERKFKFQSELVEPLNPLFCNVONDEQIOFTEDTRDKWGLDILLHCLIA
FARDDLTGDFSTSRAGAFALDISFLVSLGSAKPLMTGEGSLITLSYLGGVRA
PVNYVMGAKAGLEASVRYLASSELGSONIRVNAISAGPRTIASSAVGGLIDMIHV
EONAPLRTVTVQLEVGNTAEFLASDLASGTVQVLYVDAGEIMGW"
3047..3676
/label="ORF2"
/db_xref="PID:g142011"
/transl_table=11
/translation="MQISDYPQLNLSVPRIASVHRITGETNNVQVTVNLDGTGICKAA
VRFGLPDLMLHQISSGLIDLVDQAKGDEIDHHTNEDVGTIGLQALAKALGDRKI
VRGFNLADELALVVALDFSGRPHLSYGLQIPTVRGYTDLVREFFVALVNSQ
MTLHRIOLDGINSHHIIEATFKAFARAARMAIEVDPRAGTIPSSKGLV"
3863..>4142
/codon_start=1
/label=ORF1
/db_xref="PID:g289135"
/transl_table=11
/translation="MQSFTDTPDFQLTKNTPSVWMTSDLRKVYRTGFWMNQVSLK
GCSLTVVQGETFGLGPNAGAKTLLKLLGIIRPSGGKGLLLGQPLGDX"

```

CDS

```

BASE COUNT 1133 a 909 c 858 g 1242 t
ORIGIN

```

```

Query Match 14.3%; Score 110; DB 18; Length 4142;
Best Local Similarity 59.7%; Pred. No. 3.86e-64;
Matches 333; Conservative 1; Mismatches 224; Indels 0; Gaps 0;

```

```

Db 2327 atgtcaaatgatgaacaaattcagtcactttgtatccatccgcgataaaatggggca 2386
QY 197 ATGTTCAAGCATGACAGATTATTATGTTTGTGAGCAAAATGGTAAAGATTGTGCA 256

```

```

Db 2387 ggttagattctgattctgttagcctttgtctaacagagattgattgactggagatt 2446
QY 257 ATATTGATGTTATATCATCAATCGCATTTGCTATATGGAAGACTTACGGGACGCT 316

```

```

Db 2447 tttagcaaacatctcgctggttttgcacggccttttagatatacagcacttttctactgg 2506
QY 317 TTCTGAAACTTCACGGAAGAGGCTTCTGTGTAGCTCAAGACATAGTCTTACTCATTA 376

```

```

Db 2507 tgcagttaaagtggcggtcaaaccccaatgactgaagcgcgttagtattactttgt 2566
QY 377 CAATTGGCTCATGAGCTAAAAAATTATGACGAGAGGTGTGATGATTGTGCAACA 436

```

```

*
*

```

```

Db 2567 catatttgggtggttagagctgttcttaactacaacgttattgggagtagccaagcgg 2626
QY 437 CATATTTAGGTGGCGAATTCGCAGTTCAAAATTATAATGTGATGGGTGTTGTAAGCGA 496
Db 2627 gattagaagctagttagcttatctatagcatctgaactcgctctcaaaattctcggtta 2686
QY 497 GCTTAGAAGCAAAATGTTAAATATTTAGCATTAGCTTAGTCTCTGCTATATATTCGCGTTA 556
Db 2687 acgccatctctgctggcccccacacctggcttcttagtgcgttgggtgacattttag 2746
QY 557 ATGCAATTTCACTGGTCCAAATCCGTACATTAAAGTCAAAAGGTGTGGGTGTTTCAATA 616
Db 2747 atatgaticcatctgagcaagtagccccccttaactgcgtgactgactcagctagaag 2806
QY 617 CAATTCCTTAAAGAAATCAAGAGCGTCACCTTTAAAAACGTAAACGTTGATCAAGTAGAAG 676
Db 2807 tgggcaatacagcgcgttcttctgtagtattagctagcgggtattactgtgctcaagtc 2866
QY 677 TAGGTAAAAACAGCGGCTTACTTTRTTAAGTGACTTATCAAGTGGCGTTACAGGTGAAATA 736
Db 2867 tgtacgtatagtcaggat 2884
QY 737 TTCATGTAGATAGCGGAT 754

```

RESULT 2

```

LOCUS D90902 122056 bp DNA BCT 20-NOV-1996
DEFINITION Synecocystis sp. PCC6803 complete genome, 4/27, 402290-524345.
ACCESSION D90902
NID 91652027
KEYWORDS 50S ribosomal protein L32; ABC transporter; Mg-chelataase subunit;
Na/H antiporter; Pata protein; pleb; UDP-3-0-acyl
N-acetylglucosamine deacetylase; adenylate cyclase; adenylylsulfate
3-phosphotransferase; alkaline phosphatase; biopolymer transport
ExbB protein; c-type cytochrome synthesis protein; cell division
protein FtsH; coylaric acid a,c-diamide synthase;
enoyl-(acyl-carrier-protein) reductase; esterase; extracellular
nuclease; fmu and fmv protein; glucose dehydrogenase-B; glutamate
synthase (ferredoxin); histidinol dehydrogenase; iron-regulated
protein; lipic acid synthetase; lipophilic protein;
long-chain-fatty-acid CoA ligase; methyl-accepting chemotaxis
protein; photosystem I psam subunit; regulatory components of
sensory transduction system; regulatory protein for beta-lactamase;
sensory transduction histidine kinase; spore protein sp21; sulfur
deprivation response regulator; trna-Arg; thiamin biosynthetic
bifunctional enzyme; transposase; urease accessory protein G;
virulence associated protein B; virulence associated protein C.
Synecocystis sp. (strain:PCC6803) DNA.

```

SOURCE

```

ORGANISM Synecocystis sp.
Eubacteria; Cyanobacteria; Chroococcales; Synecocystis.

```

```

REFERENCE 1 (bases 1 to 122056)
AUTHORS Tabata,S.

```

```

TITLE Direct Submission
JOURNAL Submitted (28-JUN-1996) to the DDBJ/EMBL/GenBank databases. Satoshi

```

```

Tabata, Kazuo DNA Research Institute, Laboratory of Gene Structure
2; 1523-3, Yanauchino, Kisarazu, Chiba 292, Japan
(E-mail: tabata@kazusa.or.jp, Tel: +81-438-52-3933,
Fax: +81-438-52-3934)
2 (sites)

```

REFERENCE

```

AUTHORS Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y.,
Miyajima,N., Hirasawa,M., Sugliura,M., Sasamoto,S., Kimura,T.,
Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K.,
Okumura,S., Shimpo,S., Takeuchi,C., Wada,T., Watanabe,A.,
Yamada,M., Yasuda,M. and Tabata,S.

```

```

TITLE Sequence analysis of the genome of the unicellular cyanobacterium
Synecocystis sp. strain PCC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions
DNA Res. 3 (3), 109-136 (1996)

```

```

JOURNAL 97061201
MEDLINE

```

```

COMMENT Potential protein coding regions were assigned on the basis of
similarity search of the ORFs and GeneMark analysis.

```

FEATURES

```

Location/Qualifiers

```

```
1..122056
/organism="Synechocystis sp."
/strain="PCC6803"
1..1143
/note="ORF_ID:slr1400"
/codon_start=1
/product="regulatory components of sensory transduction
system"
/db_xref="PID:g1652028"
/transl_table=11
/transl_table="MGTASLLVADDDPNFDVLDALLADQGYELNYADSQRAINDLD
TFQPLLDDVMPCLSGVEYCMIRASRWHALPITIMTALDSKLSLANCLAAAGAD
FISKPLNGLEQLQAKMLRLKHQYDALNGNLQRESVHMVHDLRNLNLTGTO
YHSHSDAPLPPERKLLQGLQOHLVDMMLVSKQEHKRLDYQEVDDLALMG
TVLDDYEAISQKQTLVTRWENPLVAKIDPPIFQRIISLLSNAIKFSPTGAKEV
ALITEDKQVITCFIDOGGKIDELKIKIFEPYEVGTIMPINAIQIGIGLAFCKMSEA
HGAITVODNOPRGAMFSLTLPPHPPVPGWPHDQWGAADVIGK"
1263..2138
/note="ORF_ID:slr0670"
/codon_start=1
/product="hypothetical protein"
/db_xref="PID:g1652029"
/transl_table=11
/transl_table="MFKHCLICTDFDGLQRLAGVFEELSLSGITKLIIFLHTSVWED
EHIADVDESKLEAKTYLESILVGOVPPGIEVKVEYSSVYVDVNLVQLEOAIILIN
GMPVRSNLESKLFSGHSTLSLAKSTKVPVAILRPQLVSYTYVEEMALRQHLWRNLVLP
YDASAGNLYLIERLSALEKAPPGKVEYCSIFLSLEDDGMRRPPELLEIRQAEAKLAE
IKQFSPFNIITVEVRHGSPVQEIILDTAFVNDITAIASRRATLLDWTVPSTLDSI
LNRSWFPLLFFSPKG"
complement(2256..7895)
/gene="nucH"
/note="ORF_ID:sl10656"
/codon_start=1
/product="extracellular nuclease"
/db_xref="PID:g1652030"
/transl_table=11
/transl_table="MSNLIITGIIDGLPGLPKPALEYLVADIADLSMYGTEAATNG
NASTGPEFFLGSATAGDYIYVASETSGFNSFPNFTDGVANFTDGTIILFENG
SYDVGEGIDGTGRPWEHLGWAYRNGALPSPFNASEWTFSGVDALDDDAANVN
ATATPWSIADSPAGTNGLDLSTYVRIGRDLPTPTRTVAPAGSELALFVSAITYN
PDNTLFLVGGDTAIVEIDKRGQLISSMTLTAGDPADPEGLTYVNGQGFVLYVERLR
QANLTFYAGLSTRADQAVTLTGTVNGILEGVSQDPVTGGFIFVKEVNPQGFQPT
TIDGFAAGSNGPTQDSNLFDPALLGNTDIADYFALANSFVSGAIANLLVLGQ
EDGRLEVDSSGNTSELVYNADPNPLSVNQGFEVTLDDNDGLYITSESGGGDN
HPQWVYAPADYVYENNAAPVTVSVANGVNNLENATTSAILGNIIGDGLSGLNTLI
SLGASDLFIEVNGCLFLKAGYVLFDSQSSYEVTVIADDTALGNTPDATITFTLDI
LAGVNTIAPQSVYFVDDGANTITAFIDLWFGTAPAGLAIGTYGPGGLTGGDAINL
FDATGSLITGVTGSGSPSTSPFTFDNALGNPSVANLSTVGSNGAFSVIDAGEVILN
GSPGAIAGSVSNTTVGIAATGATGSENGPTGGEFTILRTGDTSTALDVTYISGDAA
NGDYAIDPTVTIPAGQSQVKISINPIDDVIESTESIVITLTDANTYDVATSSA
ATVNIENDPTIVSRQLITFMYSGADGEFIEFTNIDTSDVDMTGWFSFSDSGRVAGS
VDLSAFVQPGSVLITADATFTFANWLPATAKILGLSQNLGRSEINLYDATG
NLVDLRYGDDGFSGLTIRTONSSGWSADQLAOEIDADLWLSVDDAQNASTGDD
IGNPSIYITGVDTIVKIHIOGTGTFAPLVNNVTVIEAIVGDFQDGGDYSRNLRG
FYVEEDADGNVATSEGIFIFENGNTFIDVNVGDKVQITGVTEFFGQTDITIN
ITVTSNGTLPTAANTLPTAGTTESSOGTPOQDLEAFEGMLVKFTDITLTITEMFLD
RFNEIKLSQGRPOQTFQNDANVAGYAVREELGATITDDGLSVQNAIEINLDF
GSSFTASDIRMGDTINNLGTVLSYOHAGNASSGATWRVRSADVNTQFTKVNDRPVT
PENYVGGSLKVTGFVNLVFKTIDLSGVSTAIGQDPGADTITAEFDQTDKLVTLALAI
DADVLGLAELNDFLPGSSGNAIENLVNAGAGTYNWNYPGTQFLGDLAIVGL
IKYVSLVSGDAAILNTQAFDPNPTGENRRTPTVAOTFRDLTGTFTFVAVNHFSK
KAGSLTAGDAGNDQNDQGNFNDTRTKAAQELVTLNLTPTGNTSDYLLLGDY
NAYAQEODIKALESAGXVNLGAQFSGGNTSYVFDGOTGLTDYAFASALAAQVTGAT
EWGINADADALDNLDEGRVNIFFDGTVPYRSSDHPITVLGNLASPVPEPIANEIGV
MAENGFVFLVLPQGDVQLKFNQPPFASGTFGNWQIILEAETVINGLVNQNLGQI
GVWNAWSNNWISSTQPTNSFTLEAEVTFQIDINDDLLGDLRTVFNQGSNTLLE
GILGNYVQSGDDLTPIKYLGEAFDNNLGNWQALAEITVQGVNQVLMQNLTNQILGV
WNSSADWNWISSVFEAGSPQAIAGAGIFGVDLNAVI"
complement(8026..12255)
/note="ORF_ID:sl10654"
/codon_start=1
/product="alkaline phosphatase"
/db_xref="PID:g1652031"
/transl_table=11
/transl_table="WTINSVSLISKIGGFASSNGAEIAPDPGTRKFLVWVAGDVIEI
LDLADPTNFKIEDLALNFDGIPPGFSPVPSVAVGKAGTSPSAGIVAVSLAIRDDLNN
QEAGQVFFDATTGAFLKGVSVGLPDMWTFSPDGTKILITANEGSPNSYVDPVGSV
SVLDTGGFTNLVAQNTKDFGNAQAEQAGVRFQGFADAGNDSHLLPGLNGDASDRDG
PYVIAFGSDGTKAWTTLQENNSVALIDITATTVEAILPLGFKDHLGPNGLDASDRDG
GNIQWPIFGMYMPDSISAFVAGDQTYITANEGDARNRPSDDLLPPFDEGDFIF
IEEARLKLADPLIAPPNAEALQADPAIFNLNVTTKLGDGDFDELYVYGSRSFS
INSSGNLYDSGDEFEERLIEAIPDFNASNDNDLNRNSDKPGEVGTGIDG
RIYAVGLERIGVWYDVTTPQSPFEVQYINPRFTVDPESNLTDSGEGLIFFNAA
DPNGPLTVNENSVNTAVFENVNPTPPQLQLFHADQEAQVPAIDLPRESAV
LNAALQODINDGTAGFANTLISGSDAYIPGLFSLASDEFGVGRADLILQNELGF
QAIAFGNHEFDLTALIQDLIGNGEDNFPNFFYLSNLDFTDGNLSGLVLPDHQ
AQPNISIAATTVIDNGEKIGIVGATPTTITISSPGVTVNPPQFNGVPTPEQDAL
AELIQADVGLLAANPLKVKVLAHQMOQIAIEOELAKVRLKVDIIVAGGSNTRLIDE
NKLKRGDSAQGIYPIIKTDAGNPAVYNTDGNKYKGLRVLIDRNEGILIIIPESYDP
NRSYATVDOGVADLNAAGLIDPEIQALVDQLREVIVAKESNVGVSNYLEGSRPA
IRQETNGLNLTADANLAIKTIADNSVVISLKNGGIRDDIDIGRILVPTGCTGEPRLP
NEATDAEIGNIVKPAAGISSETDIANALSNFNGLSLITVTAELLALVHLAGASDGTN
QOGRFPQVSGFAFSPDLNRPASVDLVAIEDEAGNDLVVVRNGSELDVGPRTFMV
TLGLFANGDGYPFPPSGESVDLVLPADAPRTGLATFAPDQSEODVLAELYAANFN
SLTANQADTSPEFDVRIQNLAFRYDVTVDSTGPVDPITANEIGVVAENGFFVLLPG
GDEVQLKFNQPPASGTFRNQWQILEAETVINGLVNQNLGQIGVWNAWSNNWIS
SQTPTNSNFTLEAEVTFQIDINDDLLGDLRTVFNQGVNLSLLEGLGNYVYQSGDD
LTPPIKYLGEAFDNNLGNWQALAEITVQGVNQVLMQNLTNQIGVWNSADWNWISSN
VFAGSPQAIAGAGIFGPTTTLTADSVLV"
12582..13115
/gene="cysC"
/note="ORF_ID:slr0676"
/codon_start=1
/product="adenylsulfate 3-phosphotransferase"
/db_xref="PID:g1652032"
/transl_table=11
/transl_table="MQORGVTIWLTLGSGAKTTITHALEKLRKDSGYRLEVLGDVV
RTNLTGLGSKEDRDNIRIGFVSHLLRNGVIVLSAISPYAALRQVKKHTIGDF
LEVFNAPLACEERDKVGLYAKARSEIKGFGIDDPYEPINPDVECTDLELDE
SVGIWKQKLVLDLKYIEG"
13273..13956
/gene="exbB"
/note="ORF_ID:slr0677"
/codon_start=1
/product="biopolymer transport ExbB protein"
/db_xref="PID:g1652033"
/transl_table=11
/transl_table="MNPTELQMGKGVAMWPLLILLSVSTIIERLMFWGQVILKSSQ
TASRLTAAARDWDTAIRVAQDSRRFPIAKYLLAPLRPHDPDPFVHLAESAADDOL
ALMRGRDKILIAIALLSPGLGLGLVGLIOQLSSIQISLDGTASTAGTIGIGEALI
STAGLIIAIVSLAFYRVFQGLWFNQMRVFRKVGSELEVLYRQRWFEEAMAYDDGLTP
SPEAESLPQ"
14067..14876
/note="ORF_ID:slr0678"
/codon_start=1
/product="hypothetical protein"
/db_xref="PID:g1652034"
/transl_table=11
/transl_table="MASSPKAPKSHRKFSQSYHTRPLSLMQDNQDHQGEVRIETIPL
IDVFCIITFFILGAVLSRQQAISLDIPRAGTAPQRMFMVSLDGLGQIYVEKQP
VSQEQWVALQNHQYNPSGLIVLHASRNASYNDVQLDLTLRTVGGDRVALATLPGD
GQTPGMNPNFNPNLGLPGMTPGNAFNGANPGMNFNSNGSGGSGVYVFNSTP
LPGMDANGVSPNPGMPPGPGGAMSPDNSQSNPLPGMGTVPSPAQQ"
14912..16252
/gene="fmu and fmv"
/note="ORF_ID:slr0679"
/codon_start=1
/product="fmu and fmv protein"
/db_xref="PID:g1652035"
/transl_table=11
/transl_table="MISARQLAFLILROINRDSYDVAIDRALOKHPLSPDDRFRCT
ELIVGYVRQRITDCLIEQGLDRPGKQPPDLRRYVLQGLYLRVLDQVPAASAVNTG
VDLAKANGLKLSKVNGMLRRYRAEEQGNKILDOEKISLGEQTSFFDPLWLEFEQT
WGKAETESLCAYFNQNPISLDLRINPLKTSRVEVAOSAELENLTSMAGLPQGLRIGG
KTGAITQLPGFAEGWWTVDQASQAQWVAQILNPQPEITFDVCAAPGGKTHIAELMGD
```

...
Note: remainder of annotations omitted.

Db 102039 tggctcgggggcaaaccttgatgaccaatgcggtgagcatcattaccctgacttact 102098
 383 TGCTCATGAAGCTAAAAAATTAATGCCAAGTGGTAGCATTTGTGCAACAACATATT 442
 Db 102099 ttggcgcgctgaaggattatcccaactcaaacctgtagtgggtggtggaaggccgctcgg 102158
 443 TAGTGGCGAATTCGCAGTTCAAAATTAATGTGATGGGTGTGCTAAGGCGAGCTTAG 502

Db	102159	aaatgactgtgcgttattctggcgcagattggggcccaaaatactcggtttaatgta	102218
QY	503	AAGCAAATGTTAAATATTTAGCATTTAGACTTAGCTCGTAAATATTCGCGTTAATGCAA	562
Db	102219	tttccgtggccccatccggacccttgctcttcgcgcgtgggggcattttggatatga	102278
QY	563	TTTCAGCTGGTCCAAATCCGCTAGCATTAAGTGTAAAGTGTGGGTGGTTCGAATCAAAATTC	622

[illegible]

RESULT		3				
Locus			PMU38482	2123 bp	DNA	BCT
DEFINITION			Proteus mirabilis glutathione transferase (<i>gstB</i>) gene, complete cds.			
Accession			U38482			
			G1053074			

KEYWORDS	Proteus mirabilis strain-AF 2924.
SOURCE	Proteus mirabilis
ORGANISM	Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Proteus.
REFERENCE	1 (bases 1 to 2123)
AUTHORS	Mignogna,G., Allocati,N., Aceto,A., Piccolomini,R., Di Ilio,C., Barra,D. and Martini,F.
TITLE	The amino acid sequence of glutathione transferase from Proteus mirabilis, a prototype of a new class of enzymes
JOURNAL	Eur. J. Biochem. 211 (3), 421-425 (1993)

MEDLINE REFERENCE AUTHORS	TITLE	JOURNAL MEDLINE REFERENCE AUTHORS	TITLE	JOURNAL
931/0270	2 (bases 1 to 2123)	Perito,B., Allocati,N., Casalone,E., Masulli,M., Dragani,B., Polinelli,M., Aceto,A. and Di Ilio,C.	Molecular cloning and overexpression of a glutathione transferase gene from <i>Proteus mirabilis</i>	Biochem. J. 318 (Pt 1), 157-162 (1996)
	3 (bases 1 to 2123)	Perito,B. and Casalone,E.	Direct Submission	Submitted (13Oct-1995) Enrico Casalone, Dipartimento di Biologia

```
CDS
      complement(1211...1738)
      /note="orf2"
      /codon_start=1
      /db_xref="PID:g1053077"
      /transl_table=11
      /translation="MGHPKGCIVPEVSVGLCEDALPISDIAPNILLELETLAGGAT
LHNVDQCVAARQCOQGPIVLVKHSRAGRFRDREMLLVATAHSHWYSRPFLPDG
ERQPVGSDLTSGMLVDLLKGVELKTALEHVAAYEVMLKTKEMNEYELQLVAAAD
QMVPHTNFCACTOLD"
```

BASE COUNT	625 a	450 c	403 g	643 t	2 others
ORIGIN					
Query Match		8.6%			
Best Local Similarity		62.5%			
Pred. No.		5.34e-29;			
Matches	160; Conservative	1;	Mismatches 95;	Indels	0; Gaps 0;

dbb 10 tagagcaaacgctacgtttatatggcacaatgcatacggtccotgaagggtattcgtgttaacg 69
py 500 TAGAAGCAAAATGTTAAAATATTTCACATTAGACTTTAGGTCTCGATTAATATCCGGCTTAATG 559
dbb 70 gtactctgcaggtcccaatttgtacccttagcacgatcttggtataccaagatttcctgtaaaa 129
py 560 CAATTTGAGCTGGTCCCAATCCGTACATTAGTGCAAAAGSTGTGGGTGGTTTCAATAACAA 619
bb 130 tgyttagocattgcgaaatctgtcaccacctttacgtctgcacagtcacactacagaagtatgg 189
py 620 TTCTTAAGAAATCGAAGAGCGTGCACCTTTAAACCGCTAACGTTGATCAGTAGAAGTAG 679
bb 190 gtaatactgcagcttttatgtctgaactatctgcggggataccacgggtgaatacttc 249
py 680 GTAACAACAGCGGCTTACTTTRTAAGTGACTTATCAAGTGGCGTTACAGGTGAATAATTC 739
bb 250 acgtagacggtgggttt 265
py 740 ATGTAGATAGCGGATT 755

4	RESULTS	10111 bp	DNA	BCT	27-SEP-1996
	TITLE	Haemophilus influenzae	from bases	1805686	to 1815756 (section 161
	DEFINITION	of 163) of the complete genome.			
	ACCESSION	U32846	L42023		
	IID	g1574588			
	KEYWORDS	Haemophilus influenzae.			
	SOURCE	Haemophilus influenzae			
	ORGANISM	Eubacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;			
	REFERENCE	Haemophilus.			
		1 (bases 1 to 10111)			

CC Address all correspondence to: J. Craig Venter The Institute for
 CC Genomic Research 9712 Medical Center Dr, Rockville, MD, 20850, USA
 CC Hamilton O. Smith is with the Johns Hopkins University School of
 CC Medicine, Baltimore, MD, 21205. Joseph M. Merrick is with the State
 CC University of New York, Department of Microbiology, Buffalo, NY,
 CC 14214. Chris Fields' current address is the National Center for
 CC Genome Resources, Sante Fe, NM, 87505. All other authors are with
 CC The Institute for Genomic Research, 9712 Medical Center Dr,
 CC Rockville, MD, 20850. Predicted open reading frames were determined
 CC using GeneMark software, kindly supplied by Mark Borodovsky,
 CC Georgia Institute of Technology, Atlanta, GA, 30332. e-mail:
 CC mark@bmc.gatech.edu Submission and annotation: Owen White e-mail:
 CC white@tigr.org Curation of the genome is ongoing and comments to
 CC the authors are appreciated. Biological role information, putative
 CC identifications, sequence alignments, on-line name and sequence
 CC search capability are available at TIGR's World Wide Web site
 CC (<http://www.tigr.org/db/mdb/hidb/hidb.html>).
 FH Key Location/Qualifiers
 FT source 1..10111
 FT /organism="Haemophilus influenzae"
 FT /complement(1685...2581)
 FT /gene="H11732"
 FT /note="similar to SP:Q03155 percent identity: 28.6;
 FT identified by sequence similarity; putative. E. Koonin:
 FT weak similarity to AIDA_ECOLI, BLAST score: 88"
 FT /codon_start=1
 FT /product="adhesin (aida-I)"
 FT /transl_table=11
 FT /db_xref="PID:g1574589"
 FT /translation="MNKIFVIMNVVTQTWVYVSELTRAHTRTSATVATVATVLS
 A TVQAINDAGTFVYQSTEDIEDSATKDDNNQALKAGDTLTLKAGNLKAKLDOGG
 FT K SVTFALAKLDVKTAKVSDTLTIGGNTPAAGATPKVSTSTADGLKIAKNGDTAV
 FT H LNSTLTPDVTNTGASTSVTFSPSDIEKTRAATIKDVLNAGNWKAKVAGNGTEN
 FT V DLVAGYDNVEFITGDKNTLDVLTAKENGKTTEVKTPTKTSIKDNNKLLTGKQLKD
 FT A NTGTANNATEDDEAMA"
 FT complement(2937..4916)
 FT /gene="H11733"
 FT /note="similar to SP:P30850 percent identity: 50.8;
 FT identified by sequence similarity; putative. E. Koonin:
 FT homolog of RNB_ECOLI, BLAST score: 529"
 FT /codon_start=1
 FT /product="exoribonuclease II (RNaseII)"
 FT /db_xref="PID:g1574590"
 FT /translation="MFQDNPLLAQLKQIHSKEQVEGVVSKTDKAYGLECDKITYF
 FT I APPSMKVMHGDKIKATIEKQKQAEPEALIEPMLTRFIAKVRFNKDKLQVLVDH
 FT P SINQIPGAQAKSVKELOEGDWMVAVNLKTHPLRDRDFYATINQLICRADDELAPWW
 FT V SINQIPGAQAKSVKELOEGDWMVAVNLKTHPLRDRDFYATINQLICRADDELAPWW
 FT T TLARHQSRYPVGAEPYEMLDQKTRENTALHFVITDSESTMDDALYIEPIAQS
 FT E QTGWKLVAIADTAYIALDSQTEQBAKQRCFTNPLPGNIPMLPRELDELCSLIAN
 FT Q TRPALVCYIETDLTGNITAKPHFVSAYVQSKAKLAYNKVSDYLEQADNAWQPEMPEA
 FT I QIHLWQFTKARTQWRKTHSLFFKPEPDYAFVLAENGKVQEIKAERYRRTANOIVEAM
 FT L IANICAAQFLHEQAKTGIFNTHSGFKFLFNHNFMANLANEONQOTELABRYSVEN
 FT Y ATNLGYCOMRHDIETESDYELRLRRLYTLFAEFKSELAPHFGLGLEGYATWTSPIRK
 FT N SDMVNHLIKAVLAKOPYEKPQNDVILARQEARQNRNLVERDIADWLYCRYLADKVAS
 FT I AEFEAEVQDVMRAGRLVQLLENGASLIFIPAATLHNNKEEIQLPNDELALYIKGERTYK
 FT GDMVKVRLTEVKETRISVGEILQ"

complement(4999..5886)
 /gene="H11734"
 /note="similar to GB:M97219.1 percent identity: 75.3;
 identified by sequence similarity; putative. E. Koonin:
 homolog of FBLI_ECOLI, BLAST score: 978"
 /codon_start=1
 /product="short chain alcohol dehydrogenase homolog (envM)
 " /transl_table=11
 /db_xref="PID:g1574591"
 /translation="MRLVFLLEILVGFVQRIQFAYTQVYFANNIGKIMGFLTGKRLIV
 GLASNRSIAYGIAKSMKEQGAELAFYTLNDKLPVVEFAKEFGSDIVLPLDVATDES
 QNCPAELSKRDKFDFGFIHAIAFAPGPDGSDYVNAATREGYRIAHDISAYSFVAMAQ
 ARPYLNPAALLLSYLGAERATPNYVMCLAKASLEAATRVMAADLGKSGIRVNAIS
 GPIRTLAASGIKNFKKMLSTFEKTAALRTVTIEDVGNSSAAFLCSDLASGITGEIVHV
 AGFSITAMGELGEE"
 5966..7549
 /gene="H11735"
 /note="similar to GB:U14003.286 percent identity: 86.0;
 identified by sequence similarity; putative. E. Koonin:
 homolog of RF3_ECOLI, BLAST score: 2387"
 /codon_start=1
 /product="peptide-chain-release factor 3 (prfC)"
 /transl_table=11
 /db_xref="PID:g1574592"
 /translation="MSYPLEEVNKRRTFAIISHPDAGKTTITEKVLVLYGNAIQTAGSV
 K GKGSAAHAKSDWMEKQGISITTSVMQFPYNDCLVNLDTGHEDFSEDYRTTLTA
 V DCLMVIDSAKGVBEERTIKLMEVTRLRDTPITFMNKLDRDIRDPIELLEDVENVLKI
 R CAPITWPGCGKLFKGVYHLAKDITYLQSGQSTIOAVRVKGLNPNPELDVAVGDGL
 A QOLRELELVQGSNEFEQDAFIKGELTPVFFGTALGNFGVDHFLDGLTQWAKPKQSR
 Q ADRTVESAEKFSFGVFKIOANMDPKHRDVRVPMRVVSGYKGMKLVKVRIGKDVV
 SDALTFMAGDRAHAEAYAGDIIGLHNHGTIQIGDFTQGETLTKFTGIPNFAPELFR
 I RLKDLPLKQKQLKGLVOLSEGAQVQFVPLNNDLIIVAGVGVLOFVWVSRKLEYVW
 E AIYENVNVATARWVECADEKKFEFKKNEQNLALDGGDNLTYIAPTVMNLNAQERY
 P DVVFYKTRH"
 complement(7625..7858)
 /gene="H11736"
 /note="identified by GeneMark; putative"
 /codon_start=1
 /product="H. influenzae predicted coding region H11736".
 /transl_table=11
 /db_xref="PID:g1574594"
 /translation="MDNFIEFLGYMATFFVAASFLFKSIVHLRIVNSIGAILFVYS
 ... Note: remainder of annotations omitted.
 Query Match 7.3%; Score 56; DB 13; Length 10111;
 Best Local Similarity 58.5%; Pred. No. 1,87e-21;
 Matches 186; Conservative 1; Mismatches 131; Indels 0; Gaps 0;
 Db 5035 gaaacctgcacagtcgaacgatttcgccagtaatgccagagcgctagtcagacataa 5094
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 756 GAATCGGCTATCATGAATATTTTCCACTGTACGCCACTGTAGTACTTAATAA 697
 Db 5095 aaatcgctgagttaccacacatcttcgacagtcgacagtcgacgagcgtttt 5154
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Cp 696 GTAACGCCGTGTTTACCTACTTACTTGATCAACGTTACGTTTTTAAAGGTGCACGCTC 637
 Db 5155 ttcaagggtggaagcattttctgaagtttttaagtcgctgtagcggtgtagcggt 5214

```

Cp 636 TTCGATTTCTTAAGATTGATTGAACACCACCACACCTTTTCACCTTAATGTACGGAT 577
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 5215 tgggcctgcagagattcattcacacgaatcctcttcttgctaaatctgctgcattac 5274
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Cp 576 TGGACCAAGCTGAATTCGATTAACCGCAATATTAATCAGGACCTTAAGTCTTAATGTAATA 517
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 5275 gcgtgtgcggctcaagagagggtttgttaggcacattacgttagttaggaattgc 5334
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Cp 516 TTTAACATTGTTCTTAAGTCGCTTTAGCAACACCCATCACATTATAATTTTGAATGC 457
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 5335 gcgctcagcacctaagta 5352
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Cp 456 GAATTCGCCACCTAAATA 439
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |

RESULT 6
LOCUS ECOENV 1366 bp DNA BCT 25-APR-1994
DEFINITION Escherichia coli short chain alcohol dehydrogenase homolog (envM)
          gene, complete cds.
SSION M97219
KEYWORDS envM gene; short chain alcohol dehydrogenase.
SOURCE Escherichia coli DNA.
ORGANISM Escherichia coli
          Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
          Escherichia.
REFERENCE 1 (bases 1 to 1366)
AUTHORS Bergler H., Hogenauer, G. and Turnowsky, F.
TITLE Sequences of the envM gene and two mutated alleles in Escherichia
      coli
JOURNAL J. Gen. Microbiol. 138, 2093-2100 (1992)
MEDLINE 93123967
REFERENCE 2 (bases 1 to 1366)
AUTHORS Bergler, H.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-1992) Helmut Bergler, Institut fuer
      Mikrobiologie, Universitaet Graz, Universitaetsplatz 2, Graz,
      A-8010, Austria
FEATURES
  source
    1..1366
    /organism="Escherichia coli"
    /strain="K12, W3110"
  misc_feature
    144..152
    /note="partial sequence of boxC"
  misc_feature
    199..241
    /note="boxC"
  misc_feature
    254..342
    /note="partial sequence of boxC"
  -10_signal
    311..316
    /gene="envM"
    /note="putative"
    404..1192
    /gene="envM"
    /codon_start=1
    /db_xref="PID:q145851"
    /transl_table=11
    /translation="MGFLSGKRLVTVGASKLSIAYGIAQAMHREGAEIAFTYQNDKL
      KGRVEFAQLGSDIVLQCDVAEDASIDTMFAELGKWPFDGFGVHSIGFAPGQDLDG
      DYNAVTRREGFKIAHDISSYFVAKACRMLNPGSALLTSLYLGAEIRAIYNVMG
      LAKASLEANRYMANAMGPEGVYVNAISAGPIRTLAASGIDKDFKMLAHCEAVTPIRR
      TVTIEDVGNAAFLCSDLGASIGSEVVHVDGGSIAAMNELEK"
    680
  allele
    /gene="envM"
    /note="change in base causes diazaborine resistance (Gly
      -> Ser)"
    1125
  allele
    /gene="envM"
    /note="change in base produces ts-phenotype (Ser -> Phe)"
BASE COUNT 307 a 368 c 355 g 336 t
ORIGIN

Query Match 5.69; Score 43; DB 21; Length 1366;
Best Local Similarity 57.0%; Pred. No. 4.14e-12;

```

```

Matches 166; Conservative 1; Mismatches 124; Indels 0; Gaps 0;
Db 866 aactacaagcttatggtctgcgaagcgtctctgaagcgaacgtgcgttatggcg 925
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
QY 466 AATTATATGTAAGTGTGCTTAAGCGAGCTTAGAAGCAAAATGTTAAATATTTAGCA 525
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 926 aacgcgatgggtccggaaggtgctgtttaacgcgcattctctgtgcgcattccgtact 985
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
QY 526 TTAGACTTAGTCTCTGATATATTCGCGTTAATGCAATTCACGTGGTCCAACTCGTACA 585
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 986 ctggcggtcccggtatcaaaagacttcgcgaatgctggtcgtcgttcgcgaagcggtacc 1045
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
QY 586 TTAAGTGCAAAAGGTGTGGGTGTTTCAATACAAATCTTTAAGAAATCGAAGAGCGTGCA 645
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 1046 cggattcgcgcgtaccgttactattgaagatgtgggtgaactctgcgcattcgtctctc 1105
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
QY 646 CCCTTAAAGCGTAACGTTGATCAAGTAGAGTAGGTAAGCAACGCGCTTACTTCTTAAGT 705
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
Db 1106 gatctctgcgcgtatctcccggtgaagtgtccacggttgacggcggtttc 1156
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
QY 706 GACTTATCAAGTGGCGTTACAGGTGAATAATTCATGATAGTAGCGGATTC 756
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |

RESULT 7
LOCUS ECENVMACP 1905 bp DNA BCT 21-NOV-1996
DEFINITION E.coli envM gene.
ACCESSION X78733
NID 9587105
KEYWORDS envM reductase; enoyl-ACP reductase; envM gene.
SOURCE Escherichia coli.
ORGANISM Escherichia coli
          Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
          Escherichia.
REFERENCE 1 (bases 1 to 1905)
AUTHORS Kater, M.M., Koningsstein, G.M., Nijkamp, H.J. and Stuitje, A.R.
TITLE The use of a hybrid genetic system to study the functional
      relationship between prokaryotic and plant multi-enzyme fatty acid
      synthetase complexes
JOURNAL Plant Mol. Biol. 25 (5), 771-790 (1994)
MEDLINE 94355651
REFERENCE 2 (bases 1 to 1905)
AUTHORS Kater, M.M.
TITLE Direct Submission
JOURNAL Submitted (08-APR-1994) M.M. Kater, Institute for Molecular
      Biological Sciences, Dept of Genetics, BioCentrum Amsterdam, Vrije
      Universiteit, De Boelelaan 1087, 1081 HV Amsterdam, NETHERLANDS
FEATURES
  source
    1..1905
    /organism="Escherichia coli"
    /strain="K12, W3110"
    /clone_lib="Sau3A partial library in pUC19"
    /clone="pEN"
    /map="min 28.5 of the E. coli"
    437..442
    448..1318
    /evidence=experimental
    450..1318
    /evidence=experimental
    530..1318
    /evidence=experimental
    /gene="envM"
    /codon_start=1
    /product="enoyl-ACP reductase"
    /db_xref="PID:g587106"
    /db_xref="SWISS-PROT:P29132"
    /transl_table=11
    /translation="MGFLSGKRLVTVGASKLSIAYGIAQAMHREGAEIAFTYQNDKL
      KGRVEFAQLGSDIVLQCDVAEDASIDTMFAELGKWPFDGFGVHSIGFAPGQDLDG
      DYNAVTRREGFKIAHDISSYFVAKACRMLNPGSALLTSLYLGAEIRAIYNVMG
      LAKASLEANRYMANAMGPEGVYVNAISAGPIRTLAASGIDKDFKMLAHCEAVTPIRR
      TVTIEDVGNAAFLCSDLGASIGSEVVHVDGGSIAAMNELEK"
    1338..1356
  terminator
    /note="hairpin termination of the transcription"

```



```
BASE COUNT      422 a      519 c      494 g      470 t
ORIGIN
Query Match      5.6%; Score 43; DB 21; Length 1905;
Best Local Similarity 57.0%; Pred. No. 4.14e-12;
Matches 166; Conservative 1; Mismatches 12; Indels 0; Gaps 0;

Db 992 aactacaagcttatgcttcgcaaaagcgtctctggaagcgaacgtcgctatatggcg 1051
  || || || || || || || || || || || || || || || || || || || || ||
QY 466 AATTATAATGTAAGGTTGTTGTAAGCGAGCTTGAAGCAAAATGTTAAATATTAGCA 525
  || || || || || || || || || || || || || || || || || || || || ||
Db 1052 aacgcgatggctcggaaggtgctgttaacgccatctctctggtcgatccgtact 1111
  || || || || || || || || || || || || || || || || || || || || ||
QY 526 TTAGACTTAGTCTGATATAATTCGCTTAATGCAATTCAGTTCGTCCTCAATCCGTACA 585
  || || || || || || || || || || || || || || || || || || || || ||
Db 1112 ctggcgctccggtatcaaaagacttcgcgaataatgctggtcattgcgaagcgttacc 1171
  || || || || || || || || || || || || || || || || || || || || ||
QY 586 TTAAGTGCAGAAAGGTGGGTGTTTCAATACAAATCTTAAAGAAATCTCGAAGCGGTGCA 645
  || || || || || || || || || || || || || || || || || || || || ||
  1172 ccgattccgcgtaccgttactattgaagatgtggttaactcttcggcattcctgtgctcc 1231
  || || || || || || || || || || || || || || || || || || || || ||
  646 CCCTTAAAGCTAACGTTGATCAAGTAGAGTAGGTAACACAGCGGCTTACTTCTTAAGT 705
  || || || || || || || || || || || || || || || || || || || || ||
Db 1232 gatctctgcggtatctccggtgaagtggtccacgttgacgcgcggtttc 1282
  || || || || || || || || || || || || || || || || || || || || ||
QY 706 GACTTATCAAGTGGCGTTACAGGTGAAATATTCATGTAGATAGCGGATTC 756
  || || || || || || || || || || || || || || || || || || || || ||

RESULT      8
LOCUS      ECAGE000227 10477 bp      DNA      BCT      21-JAN-1997
DEFINITION      Escherichia coli from bases 1346941 to 1357417 (section 117 of 400)
                of the complete genome.
ACCESSION      AE000227 U00096
NID      91787543
KEYWORDS
SOURCE      Escherichia coli.
            Escherichia coli
            Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE      1 (bases 1 to 10477)
AUTHORS      Blattner,F.R., Plunkett,G., III, Mayhew,G.F., Perna,N.T and
            Glasner,F.D.
TITLE      Direct Submission
JOURNAL      Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            This sequence was determined as part of the E. coli Genome Project
            (Frederick R. Blattner, director) at the University of
            Wisconsin-Madison. Supported by award HG00301 from the NIH Human
            Genome Project. The entire sequence was independently determined
            from E. coli MG1655; overlaps with other sequence determinations
            are annotated.
FEATURES
            Location/Qualifiers
            1..10477
            /organism="Escherichia coli"
            /strain="K-12"
            /sub_strain="MG1655"
            complement(64..1269)
            /note="f401; This 401 aa orf is 25 pct identical (7 gaps)
            to 166 residues of an approx. 600 aa protein HEXA_ALT50
            SW: P488231; UUG start"
            /codon_start=1
            /db_xref="PID:g1787544"
            /transl_table=11
            /translation="MLSPIRLSPALRQDNDFLYDQAGPMEQRHITGKSHWHETOS
            STTEYDVLVPEAKVSDPFLDLVILEKETLAPFLSWLPARVLAVDLFPDQITVTR
            SOTFTAYRSLTVAQVCGVQVRLPGLPDPSTRESNRLAQITQYARQ
            LASSPSIIDNRSROHLDVNDWCVIISIIIGFIOGFIATFOAYILHPVRLP
            GLEQNTADASLFADESIRWESSVEVEKPEHTKSSFAELCOLAEILSLHPISLSLL
            EKLNTSTGNTQPNQALALCARINGSAPACFATCMSSNEYKKSITLMRKGNEINQ
            WADHRSVETATVQAIQWLTRAPDRFSAQAQFSPILLEHEKSSQIINLLVWSGLCGWINR
            LKIALGETY"

misc_feature
misc_feature
CDS
```

```
complement(755..2652)
/note="corresponds to x78733; (ECENVMACP) 1..1905"
complement(1157..2526)
/note="corresponds to M97219; (ECOENVM) 1..1366"
complement(1335..2123)
/genes="fabI"
/EC_number="1.3.1.9"
/note="f262; 100 pct identical to FABI_ECOLI SW: P29132;
CG Site No. 812; alternate name fabI"
/codon_start=1
/product="enoyl-[acyl-carrier-protein] reductase (NADH)"
/db_xref="PID:g1787545"
/transl_table=11
/translation="MGFLSGKRILVTGVASKLSIAYIAQAHREGAELAFYQNDKL
KGRVEFAQLGSDIVLQCDVAEDASIDTFAELGKVPKFDGFVHISAFPGDQJG
DVNAVTRREGFKIAHDISSYFAMAKRSMNPSCSALLTISYLAERAIINNYMG
LAKASLEANRYMANAMGEGYVNAISAGPTITLAASGIDKDFKMLAHCEAVTPIRR
TVTIEDVNSAFLCSDLSAGISGEVHVVDGFSIAAMNELELK"
complement(2491..2844)
/genes="ycjD"
/note="f117; 100 pct identical to ycjD_ECOLI SW: P45736"
/codon_start=1
/product="hypothetical 14.0 kD protein in envM-sapF
intergenic region"
/db_xref="PID:g1787546"
/transl_table=11
/translation="MIDKIKSNARDLRRNLTLQERKLWYLRSRRESDFKFRQHPVG
SYLDPAFCGSAARVVVELDGGQHDVAVDSRRTSWLESQGTVLRFWNNEIDCNEETV
LENILQELNRRSPSP"
complement(2521..3814)
/note="corresponds to U08190; (ECU08190) 1..1294"
complement(2912..3718)
/genes="sapF"
/note="f268; 100 pct identical to SAPF_ECOLI SW: P36637"
/codon_start=1
/product="peptide transport system ATP-binding protein
SapF"
/db_xref="PID:g1787547"
/transl_table=11
/translation="MIETLEVRNLSTFRYRTGWFRQTVAVKPLSFTLRGQTLLA
IIGENGSKSTLAKMAGMIPTSGELLIDHPLHFGDYFSRQIRIMIDFQSTSLN
PRORISQILDFFPLRLANTDPEQRKQIETMRMVLGLPDHVSYPHMLAPGOKQRLG
LARALIKPKVIIADEALASLDMSRSLINLMLLEQEKQISIVYVTOHGMKHS
DQVLVHQGEVRETVADVIASPLHETLRLAGHGEALTADAWKDR"
complement(3720..4712)
/genes="sapD"
/note="f330; 100 pct identical to 30 aa fragment
SAPD_ECOLI SW: P36635 but has 300 additional N-terminal
residues; 96 pct identical to 330 aa SAPD_SALTY SW:
P36636"
/codon_start=1
/product="peptide transport system ATP-binding protein
SapD"
/db_xref="PID:g1787548"
/transl_table=11
/translation="MPLLDIRNLITIEKTGDEWKAVDVRSWMLTEGEIRGLVCESSG
KSLIAKAICGVNKNRVTADRRMRDIDLLRLSAREKRLKGVNHSWMLFQEPQSL
DPSERVGRQLMQNIPAWYKGRWQRFKRAIELLHRVGIKHDKADMSFPFVELT
EGCOKVMIAIALANOPRLIADPTNSMEPTQAOIFRLITRLNQNSNTLILISHD
LQMLSDWKINLVYCGQVETAPSKELVTMPHPHYTOALITRAIPDFGSAMPKSRNLN
TLGAIFLLQLPIGCRPLGPRCPYAQRECIVTPLRTGAKNHLACHFPLNMEKE"
complement(4712..5602)
/genes="sapC"
/note="f296; 100 pct identical to GB: ECSAPABCD_2
ACCESSION: X97282; 94 pct identical to SAPC_SALTY SW:
P36669"
/codon_start=1
/product="peptide transport system permease protein SapC"
/db_xref="PID:g1787549"
/transl_table=11
/translation="MPYDSVSEKRPPTGLTAWRKFYSDASAMVGLYGCAGLAVLCI
FGWFAFYDQQLGQVQLPPSRYSYGEVFEFLGTDLDGRDVLRLSLSGAAPVGA
FVTLAATIGCLVGTAGATHGLRSVAVLNHILDTLAIPLSLLAIIVVAFGLSLH
```


Headed by:

Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishiori@nibb.ac.jp
Information operator:
Name: Hirokazu Mori
Address: NARA Institute of Science and Technology,
Ikoma, 630-01, Japan
E-mail: hmori@gcc.aist-nara.ac.jp

URL:

The Japan E. coli genome database
http://bsw3.aist-nara.ac.jp.

FEATURES
source

Location/Qualifiers
1..17297

/organism="Escherichia coli"
/note="Nucleotide position 1340563-1357859 from the
initiation site of *Thra* (0 min.); This clone is from
Kohara lambda miniset library"

/strain="K12"

/clone_lib="Kohara clone #255"

/map="28.8 min"

<1..771

/gene="acnA"

/note="ORF ID: o254-5#1; similar to [PIR Accession Number
S22375]"

/codon_start=1

/product="Aconitate hydratase (EC 4.2.1.3)"

/db_xref="PID:g1742090"

/transl_table=11

/translation="DSTVIRLSPFFDEMQATPAPVEDIHGARILAMGLSDSVTDDHSP
AGSKPDSAGRYLQGRVERKDFNSYGRNNEVMRGTFANIRIRNEMVPGVEGG
MTRHLPDSVSIYDAARYKQEQTPAVTAGKEYGSGSRDWAAGKPRLLGIRVIA
ESFERHNLGILGILPEFPQGVTRKTLGTGEBEKIDGDLQNLQPGATVPVTLIR
ADGSEVVPFCRIDTAFETYQNDGILHYVIRMLK"
complement(835..1425)

/gene="rba"

/note="ORF ID: o255#2; similar to [PIR Accession Number
A40654]"

/codon_start=1

/product="GIP cyclohydrolase II (EC 3.5.4.25)"

/db_xref="PID:g1742091"

/transl_table=11

/translation="MQLKRVAAEAKLTPWGDPLMVGFELATGHDHVALVYGDISGHT
PVLAVHSECTLDALFSLRCDGFOLEAALTOAEGRGILLYHROEGRNIGLNKI
RAYALQDGYDVYENHQLGEAADERDTLCADWFKLLGVNEVRLTNNPKKVELTE
AGINIVERPLVGRNPNNEHYLDTKAEKMGHLNK"
1595..2359

/gene="pqpB"

/note="ORF ID: o255#3; similar to [SwissProt Accession
Number P18201]"

/codon_start=1

/product="Phosphatidyglycerophosphatase B (EC 3.1.3.27)"
/db_xref="PID:g1742092"

/transl_table=11

/translation="MRSIARTAVGAALLVMPVAVVWISGNWQPGESWLLKAFWV
TETVTPGWVTHILFGLFWLCLFRKAAFLFAALAAIIVGGVKSWKQVQE
PRPVLWLEKTHHPVDEFTLKRAGNVLKVEOLAEKNIPOYLRSWQKGTGPAF
SGHTMFAASWALLAVGLLWPRBRTLIAIIVYATGVMSRLLGMHWPDLVWATLI
SWALVAVWLAQRICGLPTPAEENRETAQREQS"

2508..2816

/note="ORF ID: o255#4; similar to [SwissProt Accession
Number P44129]"

/codon_start=1

/db_xref="PID:g1742093"

/transl_table=11

/translation="MKYLLIFLLVLAIFVISTLQANDQQTFFNYLLAQGYRISTIL
LAYLFAAGFAIGLWGLFWRVRLVRLAERKIKKLENQLSPATDVAVVPSSAAKE
"

2823..3992

/gene="yciM"

/note="ORF ID: o255#5; similar to [SwissProt Accession
Number P45576]"

/codon_start=1
/db_xref="PID:g1742094"
/transl_table=11
/translation="MLELLFLLPVAAYGWYMGRRSAQOONKDEANRLSRDYVAGVN
FLSNQOKAVDLELMDLKEDTGVAAHLTGNLFRSGEVDRAIRIHOTLMEASLT
YEORLLAQGLGRDYMAGLIDRAEDMFNQLTDETRIGALQQLLOIYQATSEWOKA
IDVAFRLVKGKDKORVIAHFYICELAQHNASDDDLRAMELTKKRAADKNSARVSI
MMGRVYAKGEYAKVESLORVISOQRELVSLEMLQTCYQQLGKTAWEAFLOKAV
EENTGADAEMLADIIEARDGSEAAQVITRQLQRHPTMVFVHKLMDYHLNEAPEGRA
KESLMVLDRMVGKVRSPRYRCQCGGTAVTLYWHGCPSCRAWSTIKPIRGIDGL"
4186..4923
/gene="pyrF"
/note="ORF ID: o255#6; similar to [SwissProt Accession
Number P08244]"
/codon_start=1
/product="Orotidine 5'-phosphate decarboxylase (EC
4.1.1.23) (OMP decarboxylase)."
/db_xref="PID:g1742095"
/transl_table=11
/translation="MTLASSSSRAVNTSPVVVALDYHNRRDALAFVDKIDPRCRLK
VGKEMFTLFGQFVRELQGRFDIFDLKFDHDPNTAAHAAADLGVWVNVNHSAG
GARMATAAREALVFPFGDAPLLIATVLTSMESADLVLDGMTLSPADAEALATOK
CGLDGVGCSAQAEAVREFKOVFGQEFKLVTPGIRPQSGEAGDQRRIMTPSQALSAGVDYM
VIGRPVTSQSDPAQTLKAINASLQRSA"
4923..5249
/gene="yciH"
/note="ORF ID: o255#7; similar to [SwissProt Accession
Number P08245]"
/codon_start=1
/db_xref="PID:g1742096"
/transl_table=11
/translation="MSDSNRLVYTERGRIDEKPAAPVRPKGDGVVRIQRTSGRKG
KGVCLITGVLDLDAELTKLAELKKKCGGAVKGVIEIQGDKKLLKLLLEAKGMK
VKLAGC"
complement(5375..5593)
/gene="osmB"
/note="ORF ID: o255#8; similar to [PIR Accession Number
A32255]"
/codon_start=1
/product="Lipoprotein osmB precursor, osmotically
inducible"
/db_xref="PID:g1742097"
/transl_table=11
/translation="MEVTSKMTAAVLATLAMSLSACSNSKRDENTAIAGAGAGALG
GAYLDTGSLTGLTGAAVGVGHQVQVK"
complement(5863..6213)
/note="ORF ID: o255#9; similar to [SwissProt Accession
Number P42902]"
/codon_start=1
/product="Putative *aga* operon transcriptional repressor."
/db_xref="PID:g1742098"
/transl_table=11
/translation="MLGGVYQKKSVMGVLPTROCIOQVHFSKAFIGIDGWPETGFT
GRDMRTDVVNAVLEKECEAIVLTDSSKFGAVHSYSGIPVERFNRVITDSKIRASDL
HLEHSLTIHVVDI"
complement(7026..9011)
/gene="MFC02B10.18C"
/note="ORF ID: o255#12; similar to [SwissProt Accession
Number Q11024]"
/codon_start=1
/db_xref="PID:g1742099"
/transl_table=11
/translation="MKTVRESTITLYNFGSHNPNYWRLTSSDVLRFSTETTTPDRTL
OLSAQAARIREMTVITSSLMMSLTVDESOLSVHLVGRKINKREWAGNASAWHDTPAV
ARDLSHGLSFAEQVVSSEAIVLDSRGNIOVLRCEDTGLKEDHDVIGQSVFKLF
MSRREAAASRRNRRVFRSGNAYVELWIPCTCKQRFLFPFNKVFHSGSGNKFILIC
SGDITERRAORERILANTDSITGPNRNAMQDLIDHAINHADNKNVGVYLDLN
FKKYNDVAGHLFGDQLLRDYSLATLSCHEDVLAARPGGDFLVLAGNTSQSALEAMA
SRILTRRLPRIGLIEVYITSCVGSIALSPHSGSDTAIIRHADTAMTAKEGRGOF
CVFTPEMNQRVFEYLWLDLTNLKRGALSDQVILVIHQPKITWRGEVRSLEALVWQSPER
GLIPLDFIYSIAEESGLVPLGRWVILDVVQVAKWRKGNLNRVAVNIARQADOT
IFTALKVQLQENFEPYCPIDVELTESLIEDELALSVIOQFSQGAQVHLDGFGTY
SSLSQARFPIDAIKQDVFVRDIHKQVPVQSLSVRAIVAVAAQALNLQVIAEGVESAKE

Headed by:

```

/seqs/prot
/notes="ORF_ID:o255#5; similar to [SwissProt Accession
Number P45576]"
/codon_start=1

```

CDS

```

/db_xref="PID:g1742094"
/transl_table=11
/translation="MLELLFLLPVAAAYGWMRRSAQNKQDEANRLSRDYVAGVN
FLSNQDQKAVDLFLDKEDGTVEAHLTLGNLFRSCEVDRAIRIHOIYMESASLT
YEORLAIQGLGNDYNAAGLYDRAEDFNQLTDEDFRIGALQQLLOLYQATSEWOKA
IDVARELYKLGDKQKQVEIAHYCEALQHMASDLDLDRMTLLKGAADKNSARVSI
MMGRVFMAGEYAKAVESQORVISOQRELYSETLMQTCYQOLGKTAWEAFLOQRA
EENTGADALMLADIIEARDGSEAAQVYITRQORHPTMRVFKMLMDYHLNPAEGRRA
KESMLVLRDWMVGKVRKSPRYRCQKCGFTAYTLWHPCSCRAWSTIKPIRGILDGL"
4186..4923
/gene="pvrF"
/notes="ORF_ID:0255#6; similar to [SwissProt Accession
Number P08244]"
/codon_start=1
/product="Orotidine 5'-phosphate decarboxylase (EC
4.1.1.23) (OMP decarboxylase)."
```

CDS

```

/transl_table=11
/translation="MLTATSSSRVNTSPVVVALDYHNRDDALAFVDFKIDPRDRLK
VGKEMFTLFGPQVRELEQGFDFDLKHFDPNTAAHAAVAAADLGVWVNVHASG
GARMTAAAREALVPFGKADPLIAVAVLTASMEASDLVGLMTLSPADYAEIRLAALTQK
CGLDVVVCSAQEAQVRFQVFGQEFKLVTPGIRPQSGEAGDQRRIMTPEQALSAGVDM
VIGRPVTSQVDPAQLKAINASLQRSA"
4923..5249
/gene="yciH"
/notes="ORF_ID:0255#7; similar to [SwissProt Accession
Number P08245]"
/codon_start=1
/db_xref="PID:g1742096"
/transl_table=11
/translation="MSDSNRLVYSTETGRIDBPKAAPVRPKDGVVVRQRTSGRKG
KGVLITGVLDLDAELTKLAELKRCGCGGAVKDGVIQDGKDLLSLLEAKGMK
VKLAGG"
complement(5375..5593)
/gene="osmB"
/notes="ORF_ID:0255#8; similar to [PIR Accession Number
A32255]"
/codon_start=1
/product="Lipoprotein osmB precursor, osmotically
inducible"
```

CDS

```

/db_xref="PID:g1742097"
/transl_table=11
/translation="MFVTSKKMTAAVLAITLAMSLSACSNSKRDNTAIGAGAGALG
GAVLTDGSLTGLTGGAAVGVGIGHQVGK"
complement(5863..6213)
/notes="ORF_ID:0255#9; similar to [SwissProt Accession
Number P42902]"
/codon_start=1
/product="Putative aga operon transcriptional repressor."
```

CDS

```

/db_xref="PID:g1742098"
/transl_table=11
/translation="MLGGVYQKKSVMVGLTRQCIQVHFSAFIGIDGWOPETGFT
GROMFTDVNNAVLRECEALVITDSKFGAVHSYSGIPVERNRVITDSKTRASDLM
HLHSKLTIRHVDI"
complement(7026..9011)
/gene="MTCY02B10.18C"
/notes="ORF_ID:0255#12; similar to [SwissProt Accession
Number Q11024]"
/codon_start=1
/db_xref="PID:g1742099"
/transl_table=11
/translation="MKTVRSTTIYNFLGSHNPYWRLTSSDVLRFSTTTETTPDRTL
QLSQAARIREMTVITSSLMSTLTVDESLSVHLVGRKINKREWAGNASAWHDTFPAV
ARDLSHGLSFAEQVVSSEHSAIVILDSRGNIORENLCEDYTGKHEHDVIGOSVFKLF
MSRREAAASRNRRFRSNAVEVLWIFTCKGQRLFLFRNKFVHSVSGSKNEIFLIC
SGDITERRAQERLRNLANTDSITGLPNRNAMDLIDHAINHADNNKVGCVVLDLIDN
FKVNDYAGHLFGDLQRLVSLAISLEHDQVLRPQGFDEFVLASNTSSQSALEAMA
SRILTRLPFRGLIENVYSCSVGALSPEHSDSTAIRHADTATVAKEGRGQF
CVFTPENNQVFEVYLDNLRKALENDOLV IYOPKITWRGEVRSLEALVRQSPER
GLIPLDFISYAESGLIYPLGRVILDVVRQVAKWRDGINLRVAVNTSARGLAQDT
IFTALKQVLELNFECIPDVELTESLIENDELAUSVIOQFSQAGVHLDDFGY
SSUSQARFPIDAIKLDQVFDIHKQPVSVQSLVRAIVAQAQNLQVIAEGVESAKE
DAFLTNGINGNERQGLFPAKMPAVAFERWYKRYLKRA"
```

CDS

```

complement(9247..11181)
/gene="rnb"
/notes="ORF_ID:0255#13; similar to [SwissProt Accession
Number P30850]"
/codon_start=1
/product="Exoribonuclease II (EC 3.1.13.1) (Ribonuclease
II) (RNase II)."
```

Note: remainder of annotations omitted.

```

Query Match      5.6%; Score 43; DB 20; Length 17297;
Best Local Similarity 57.0%; Pred. No. 4.14e-12;
Matches 166; Conservative 1; Mismatches 124; Indels 0; Gaps 0;

Db 12555 gaaacgcgcgtcaacgtggaccacttcaccggagagataccggcagagatcgggacacag 12614
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 756 GAATCCGCTATCATCAATAATTTTACCTGTAAACGCCACTTGATAAGTCACCTTAAYAA 697
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12615 gaatccgcagagttaccacacatcttcaatagtaacgtacgacgaatcgggtaacggc 12674
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 696 GTAAGCCGCTGTTTACCTACTTCTACTTGATCAACGTTACGTTTAAAGGTGCACGCTC 637
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12675 ttgcgaatgagccagcattcttgcgaagtctttgataccgagggccgcagagtacgat 12734
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 636 TTCGATTTCTTTAAGAATTGTATTGAACACCACCCACACTTTTGCACCTTAATGTACGGAT 577
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12735 cggaccacagagatg9cggttaacacgcacacaccttcggacccatcgcttgcgcata 12794
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 576 TGGACCAAGTGAATTTGCAATTAACGCGAATATTATCAGGACCTAAGTCTAATGCTAAATA 517
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12795 ggcacgttcgctccagacgacgtttgcccagaccataacgcttgtagtt 12845
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 516 TTTAACATTGCTTCTAGCTCGCTTTAGCAACACCCATCACATTATAATT 466
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

RESULT 11
LOCUS      D90767      17769 bp      DNA      BCT      17-DEC-1996
DEFINITION E.coli genomic DNA, Kohara clone #256(29.0-29.4 min.).
ACCESSION  D90767
NID         g1742107
KEYWORDS   Complete and shotgun sequencing; MTCY02B10.18C; fabI, envM; rnb;
           sapB; sapF; yjcJ; ycjD.
SOURCE     Escherichia coli (strain:K12) DNA, clone_lib:Kohara lambda miniset
           library Clone:Kohara clone #256.
ORGANISM   Escherichia coli
           Eubacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Escherichia.
REFERENCE  1 (bases 1 to 17769)
AUTHORS    Mori, H.
DIRECT SUBMISSION
TITLE      Submitted (14-DEC-1996) to the DDBJ/EMBL/GenBank databases.
JOURNAL    Hirotsada Mori, NARA Institute of Science and Technology, Res. &
           Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01,
           Japan (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
           Fax:81-7437-2-5669)
           2 (sites)
           Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiuchi, T.,
           Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K.,
           Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M.,
           Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H.,
           Motomura, K., Nakamura, Y., Nashimoto, H., Nishio, Y., Oshima, T.,
           Saico, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
           Yamamoto, Y. and Yano, M.
           The systematic sequencing of the Escherichia coli genome in Japan
           Unpublished (1996)
COMMENT    Collaboration Information:
Project:   The Japan E.coli genome DNA sequencing project
Group:     The Japan E.coli genome DNA sequencing group
```

Members: (1995.4 - 1996.3)
 Alba.H., Baba.T., Fujita.K., Hayashi.K., Hayashi.K., Honjo.A.,
 Horiuchi.T., Ikemoto.K., Inada.T., Isono.K., Isono.S.,
 Itoh.T., Kanai.K., Kasai.H., Kashimoto.K., Kim.S.,
 Kimura.S., Kitagawa.M., Kitakawa.M., Makino.K.,
 Masuda.S., Miki.T., Mizobuchi.K., Mori.H., Motomura.K.,
 Nakamura.Y., Nashimoto.H., Nishio.Y., Oshima.T., Saito.N.,
 Sampei.G., Seki.Y., Tagami.H., Takemoto.K., Wada.C.,
 Yamamoto.Y. and Yano.M.
 Headed by:
 Name: Takashi Horiuchi
 Address: National Institute of Basic Biology, Okazaki, 444, Japan
 E-mail: kishorienibb.ac.jp
 Information operator:
 Name: Hirotada Mori
 Address: NARA Institute of Science and Technology,
 Ikoma, 630-01, Japan
 E-mail: hmorigtc.aist-nara.ac.jp
 URL:
 The Japan E. coli genome database
 http://bsw3.aist-nara.ac.jp/
 Location/Qualifiers
 1..17769
 /organism="Escherichia coli"
 /strain="K12"
 /note="Nucleotide position 1347168-1364936 from the
 initiation site of ThrA (0 min.); This clone is from
 Kohara lambda miniset library"
 /clone="Kohara clone #256"
 /clone_lib="Kohara lambda miniset library"
 /map="29.0 min"
 /complement(421..2406)
 /gene="MTCY02B10.18C"
 /note="ORF_ID:0255#12; similar to [SwissProt Accession
 Number Q11024]"
 /codon_start=1
 /db_xref="PID:g1742108"
 /transl_table=11
 /translation="MKTVESTLYNFGSHPNYWRLETSSDVLRFSTTETPEPRTL
 QLSAQARIREMTVITSSLMSTVDESLSVHLVGRKINKREWAGNAGSWHDTPAV
 ARDLSGLFAGVYSEASHAIVILDSRGNIQRENFRLCEDYGLKEDHVGIGSVFKLF
 MSRRRAASRRNRNFRFSSNAVEVELWPTCKGORLEFRNKFVHSGSKNEIFLIC
 SGTDTTEERACERLILANTVSTIGLPNRNMQDLIDRAINHANNKRVGVYLDLON
 FKYNDAIGHLEGLQDLRVSLAISLCLEHQLVLPARGGDEFVLASNTSQSALEAMA
 SKILTKRPFRLGILEVTSVGLASPEHGSSTAIRHADTAMTAKEGGRQGF
 GLIPDLFISYAEESGLIPLGRWVLDVROVAKWRDKINLRVAVNTSARQLADOT
 CVFTPEMROVFYELWLDNLRKALENDQVLIHYQPKITWRGEVRSLEALVRWQSPER
 IFTALKOVLQELNFEVCPIDVELTESCLLENDELALSIVQOFSQLGAQVHLDLFGTY
 SLSOLARPEIDAIKLQVDFVRIHKQPVYSQSLVRAIVAQAALQVIAEGVESAKE
 DAFLLKNGINERQGLFAPKMPAFAFERWIKRYLKRA"
 /complement(2642..4576)
 /gene="rnb"
 /note="ORF_ID:0255#13; similar to [SwissProt Accession
 Number P30850]"
 /codon_start=1
 /product="Exoribonuclease II (EC 3.1.13.1) (Ribonuclease
 II) (Rnase II)."
 /db_xref="PID:g1742108"
 /transl_table=11
 /translation="MQDNPLLAQLKQOLHSOTPRAEGVVKATEKGFGFLEVDKQSY
 FIPPPQMKVMGDRILVIVHSEKRESAPPELYEPTLTVRGVKQKNDRALVDP
 HPLLDKALPCRAAGLNHEFKGDNAEMRRHPDGKDRFSFAETLQITTFGDDHFPV
 WWTVLARNLEKEADPGVATEMDDEGLVREDTALDFVITDSASTEDMDLAKFALP
 DKLQILVIAADPTAWIAGSKLDAKIRAFNTLPGFNPMLEPRLSDDLCSLRAN
 EYRVLACMTLSADGTEDNTEFFAATIESKALVYQVSDWLENTGDWQSESAIA
 EQVRLAQICQRRGEWRNHALVFKDRPDYRFLILGKEGLVDIVAEPRRIARIVEEA
 MTAAACARLVRDLKGLGIVNVHGFDPANADALALLKHLGLHVAEEVLTLDGFC
 KLRRLDAQPTGFLDSRIRRFQSAFISTEPGPHGLGLEATATWSPIRKQDGMNIN
 RLKLVKIGETATRFQDEITVQMAERLRNRAEDVGDVYARFLKDKAGTDTTFAA
 EIVDISRGMRVLDVNDGAIAPAPFLHVAERDELVCQENGTVOIKGETVYKVTDVI
 DVTIAEVRMETSIIARVA"
 /complement(5914..6699)
 /gene="fabI, envM"

URS
source

CDS

CDS

CDS

/note="ORF_ID:0255#16; similar to [SwissProt Accession
 Number P29132]; Author-given protein sequence is in
 conflict with the conceptual translation"
 /codon_start=1
 /product="Enoyl-[acyl-carrier-protein] reductase (NADH)
 (EC 1.3.1.9) (NADH-dependent enoyl-ACP reductase)."
 /db_xref="PID:g1742110"
 /transl_table=11
 /translation="GFLSKRLLVTVGASKLSIAYGIAQAMHREGAELAFYQNDKLLK
 GRVEFAQLGSDIVLQCDVAEDASIDTMFAELGKVPKFDGFGVHSIGFAGDQDGD
 VYNAVTRGFKIAHDISYFVAMAKASMLNPGSALTLSYLAERAIPNYVMGL
 AKASLEANYMANAMGEGVRVNAISAGPIRTLAASGIKDFRKLHAECVATPIRRT
 VTIEDVGSAAFLCDLSDGISGVEVHVDDGGFSIAAMNELELK"
 /complement(7070..7423)
 /gene="ycjD"
 /note="ORF_ID:0255#17; similar to [SwissProt Accession
 Number P45736]"
 /codon_start=1
 /db_xref="PID:g1742111"
 /transl_table=11
 /translation="MIDKISNARDLRNLTQERKLMRYLBSRFSDFKFRQHPVG
 SYILDFACSAVVVELDGGQHDLAVIDSRITSWESQGVTVLRFWNEIDCNETV
 LENILQELNRRSPSP"
 /complement(7491..8297)
 /gene="sapF"
 /note="ORF_ID:0255#18; similar to [SwissProt Accession
 Number P36637]"
 /codon_start=1
 /product="Peptide transport system ATP-binding protein
 SapF."
 /db_xref="PID:g1742112"
 /transl_table=11
 /translation="MIEITLLEVRNLSTKTVRYTGWFRQTVFAVKPLFTLEEGOTLA
 IIGENGSKSLAKMLAGNIETPSGELLIDHPLHFGYDSFSRQIRIFQDPSTSLN
 PQRIQSILDFPLKMLNLTDEPQRKQIETRMHVVLLPDDHVSYPHMLAQPKQRLG
 LARALILPKVILIADEALASLMRSQILNMLELQKGISYIVYTHIGMMKHIS
 DQVLYMHQGEVVERGSTADVLASPLHETLKRIAGHFGAETADAWRKDR"
 /complement(8299..9291)
 /gene="sapB"
 /note="ORF_ID:0255#19; similar to [SwissProt Accession
 Number P36635]"
 /codon_start=1
 /product="Peptide transport system ATP-binding protein
 SapD (fragment)."
 /db_xref="PID:g1742113"
 /transl_table=11
 /translation="MPLDILNLTIEPTGDEWVAVDRVSMTLTEGEIRGLVGSQS
 GKSLIAKICGVNKNDRVTADRMFDDIDLRLSARERKLVGHNVSMIFQEPQSC
 DFERVGRQMLQNTPAWYTKGRWQRFGRKRRATELLHRYGIDKHDMNTSPVELT
 EGECQVMIAIALANQRLIADEPTNSMEPTQAIQFRLTLRNLQNSNTTLLISHD
 LQMLQWADKINLVCGQTVETAPSKELVTPHPTQALIRAFDGSAMPKSRNL
 TLPGLAIPLEOLPGICRLGPRCPVAQRCIVTPRLTGAKNHLVACHFFLNMEKE"
 /complement(9291..10181)
 /note="ORF_ID:0255#20; similar to [SwissProt Accession
 Number P36669]"
 /codon_start=1
 /product="Peptide transport system permease protein SapC."
 /db_xref="PID:g1742114"
 /transl_table=11
 /translation="MPYDSVSEKRPPTLTATMRKFFSDASAMVGLYGCAGLAVLCI
 FGGFAPYGDIDQQLGVLQPPSWRSRGEVFFGLDGLDRVLSRLSGAAPTGGGA
 FVFTLAATICGLVGTAGATHGRSALVNIHLDTLLAIPSLALAIIVVAFAGPSLSH
 AMVAVLALPLMRVRSIYSMVHDELEKRYTAARLDGASTLNIILFWAIPNTAGLVT
 EITRALSMIADIAAGFLDGAQLPSPGWAMGLDGLLELIVVAPWTVMVLPQAAIMIS
 VLLVNLGDGVRRAIAGVE"
 /complement(10168..11133)
 /note="ORF_ID:0257#1; similar to [SwissProt Accession
 Number P36668]"
 /codon_start=1
 /product="Peptide transport system permease protein SapB."
 /db_xref="PID:g1742115"
 /transl_table=11
 /translation="MIITFLRRILLIVTFLFLTFVGFSLSYFTPHAPLAGASLWNAW


```
RL Oncological Science, University of Utah, School of Med. Rm5C334,
RL USA, UT 84132, USA
FH Key Location/Qualifiers
FT source 1..354
FT /organism="Oxytricha fallax"
FT /strain="9d1"
FT /transposon="TBE1"
FT /note="this is a bulk sequence that was generated from a
FT PCR product that represents many transposon templates"
FT CDS
FT <1..>354
FT /codon_start=1
FT /product="57kd zinc finger/protein chimera"
FT /db_xref="PID:g1881676"
FT /translation="HTRDLXKHLKAHKKXXEXXXXXLXLLKXKREXXXXXXX
FT AXEXXVXXNRNXLXSEXTKIMIKIQYKKIPVLAQIDLTSLQSYLLEDSFDKKVIX
FT HQYAYKYPFNLYLX"
FT Sequence 354 BP; 106 A; 42 C; 41 G; 54 T; 111 other;
FT
FT Query Match 3.9%; Score 30; DB 8; Length 354;
FT Best Local Similarity 26.8%; Pred. No. 1.26e-03;
FT Matches 38; Conservative 53; Mismatches 49; Indels 2; Gaps 1;
FT
FT Db 59 arrakhawgmhrhsarytraagytdctcaahaaagaaragdgrgaryhhyhdwarg 118
FT QY 137 AAGAGCTTGAATAATTTAGAACAAATTAATCAACGAGCGCACTTATATCAAAATTG 196
FT
FT Db 119 wbydggwgykaayaagcwagmarwyaswtrygtaargawdkrcgcacacwvytrr 178
FT QY 197 ATGTTCAAGCGATGAAG--AGGTTATTATGTTTGGACCAATTTGTTAAAGATGTTGG 254
FT
FT Db 179 cmtcdgagrtmacdaarathat 200
FT QY 255 CAATATTGATGGTGATATCAT 276
FT
FT RESULT 14
FT LOCUS OFU89259 354 bp DNA INV 14-MAR-1997
FT DEFINITION Oxytricha fallax 57kd zinc finger/protein chimera gene, partial
FT cds
FT ACCESSION U89259
FT NID g1881675
FT KEYWORDS
FT SOURCE Oxytricha fallax.
FT ORGANISM Oxytricha fallax
FT Eukaryotae; mitochondrial eukaryotes; Alveolata; Ciliophora;
FT hypotrichs; Stichotrichida; Oxytricha.
FT Doak,T.G., Doerder,F.P., Jahn,C.L. and Herrick,G.
FT 1 (bases 1 to 354)
FT A proposed superfamily of transposase genes: transposon-like
FT elements in ciliated protozoa and a common 'D35E' motif
FT Proc. Natl. Acad. Sci. U.S.A. 91 (3), 942-946 (1994)
FT 94134747
FT 2 (bases 1 to 354)
FT Witherspoon,D.J., Doak,T.G., Williams,K., Seger,J. and Herrick,G.
FT Selection on the protein-coding genes of the TBE1 family of
FT transposable elements in the ciliates Oxytricha fallax and O.
FT trifallax
FT Unpublished
FT 3 (bases 1 to 354)
FT Doak,T.G., Williams,K., Witherspoon,D.J. and Herrick,G.
FT Direct Submission
FT Submitted (11-FEB-1997) Oncological Science, University of Utah,
FT School of Med. Rm5C334, USA, UT 84132, USA
FT Location/Qualifiers
FT 1..354
FT /organism="Oxytricha fallax"
FT /strain="9d1"
FT /transposon="TBE1"
FT /note="this is a bulk sequence that was generated from a
```

```
CDS
PCR product that represents many transposon templates"
<1..>354
/codon_start=1
/product="57kd zinc finger/protein chimera"
/db_xref="PID:g1881676"
/transl_table=6
/translation="HTRDLXKHLKAHKKXXEXXXXXLXLLKXKREXXXXXXX
OAXEXXVXXNRNXLXSEXTKIMIKIQYKKIPVLAQIDLTSLQSYLLEDSFDKKVI
XHQYAYKYPFNLYLX"
BASE COUNT 106 a 42 c 41 g 54 t 111 others
ORIGIN
Query Match 3.9%; Score 30; DB 111; Length 354;
Best Local Similarity 26.8%; Pred. No. 1.26e-03;
Matches 38; Conservative 53; Mismatches 49; Indels 2; Gaps 1;
Db 59 arrakhawgmhrhsarytraagytdctcaahaaagaaragdgrgaryhhyhdwarg 118
QY 137 AAGAGCTTGAATAATTTAGAACAAATTAATCAACGAGCGCACTTATATCAAAATTG 196
Db 119 wbydggwgykaayaagcwagmarwyaswtrygtaargawdkrcgcacacwvytrr 178
QY 197 ATGTTCAAGCGATGAAG--AGGTTATTATGTTTGGACCAATTTGTTAAAGATGTTGG 254
Db 179 cmtcdgagrtmacdaarathat 200
QY 255 CAATATTGATGGTGATATCAT 276
RESULT 15
LOCUS MYCSEQB 1405 bp DNA BCT 17-DEC-1992
DEFINITION Mycoplasma bovine group 7 gene sequence.
ACCESSION M96587
NID g150202
KEYWORDS
SOURCE Mycoplasma bovine group 7 (strain R2222) DNA.
ORGANISM Mycoplasma bovine group 7
Eubacteria; Firmicutes; Low G+C gram-positive bacteria; Mycoplasmas
and walled relatives; Mycoplasmatales; Mycoplasmataceae;
Mycoplasma.
REFERENCE 1 (bases 1 to 1405)
AUTHORS Taylor,T.K., Bashiruddin,J.B. and Gould,A.R.
TITLE Relationships between members of the Mycoplasma mycoides Cluster as
shown by DNA probes and sequence analysis
JOURNAL Int. J. Syst. Bacteriol. 42, 593-601 (1992)
MEDLINE 93002315
FEATURES Location/Qualifiers
source
1..1405
/organism="Mycoplasma bovine group 7"
/strain="R2222"
BASE COUNT 553 a 158 c 244 g 450 t
ORIGIN
Query Match 3.9%; Score 30; DB 25; Length 1405;
Best Local Similarity 72.2%; Pred. No. 1.26e-03;
Matches 57; Conservative 0; Mismatches 21; Indels 1; Gaps 1;
Db 1178 aattattagacagctaaagaattaagagcatttagtgtaagttgctgtgaatat 1237
QY 148 AAATTATTAGAACAAATTAATCAACGAGCGCACTTATATCAAAATTGTTCAAGC 207
Db 1238 gatgaagaa-ttattaatg 1255
QY 208 GATGAAGAGGTTATTATG 226
Search completed: Fri Dec 5 07:57:15 1997
Job time : 767 secs.
```

MORSE

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
on: Fri Dec 5 07:59:43 1997; MasPar time 484.49 seconds
ular output not generated.

Title: >US-08-790-043A-2
Description: 771 1 AUGTTAACTTGAACAA.....GATTCACCAATTAATAA 771
N.A. Sequence: 771 TACAATTTAGACTTTTGT.....CTAAGTGCGTTAAATTAT
Comp:

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 707517 seqs, 256659390 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: EST-STS
1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98
99:EST99
EST-STS-TWO
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134
135:EST135 136:EST136 137:EST137 138:EST138 139:EST139
140:EST140 141:EST141 142:EST142 143:EST143 144:EST144
145:EST145 146:EST146 147:EST147 148:EST148 149:EST149
150:EST150 151:EST151 152:EST152 153:EST153 154:EST154
155:EST155 156:EST156 157:EST157 158:EST158 159:EST159
160:EST160 161:EST161 162:EST162 163:EST163 164:EST164
165:EST165 166:EST166 167:EST167 168:EST168 169:EST169
170:EST170 171:EST171 172:EST172 173:EST173 174:EST174

Database:

175:EST175 176:EST176 177:EST177 178:EST178 179:EST179
180:EST180 181:EST181 182:EST182 183:EST183 184:EST184
185:EST185 186:EST186 187:EST187 188:EST188 189:EST189
190:EST190 191:EST191 192:EST192 193:EST193 194:EST194
195:EST195 196:EST196 197:EST197 198:EST198

Statistics: Mean 10.574; Variance 2.087; scale 5.067

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description	Pred. No.
1	23	3.0	337	76	R95589	SMNHADA016014T3 Schis	3.49e-05
2	23	3.0	439	159	AA039422	zk39404.r1 Soares pre	3.49e-05
3	22	2.9	179	142	N86668	J9617F Fetal heart, L	5.54e-04
4	22	2.9	249	81	T35412	EST84470 Homo sapiens	5.54e-04
5	22	2.9	424	186	AA138421	mq88509.r1 Strata gene	5.54e-04
6	22	2.9	435	170	W57951	zd21e07.r1 Soares fet	5.54e-04
7	22	2.9	516	171	W67124	zd39a11.r1 Soares fet	5.54e-04
8	22	2.9	604	113	W02311	za08407.r1 Soares mel	5.54e-04
9	21	2.7	132	72	H85028	ys88901.sl Homo sapie	7.94e-03
10	21	2.7	172	17	T47205	ys33907.r1 Homo sapie	7.94e-03
11	21	2.7	210	182	AA092855	119883.seq.F Fetal he	7.94e-03
12	21	2.7	214	48	HUM003B01A	Human fetal brain cDN	7.94e-03
13	21	2.7	233	48	HUM238E03B	Human aorta cDNA 5'-e	7.94e-03
14	21	2.7	284	24	R44860	yg25a12.sl Homo sapie	7.94e-03
15	21	2.7	294	10	T93954	yg6h07.r1 Homo sapie	7.94e-03
16	21	2.7	313	48	HUM322D12B	Human aorta cDNA 5'-e	7.94e-03
17	21	2.7	335	77	CELK107GSR	C.elegans cDNA clone	7.94e-03
18	21	2.7	337	36	R06193	ye94e04.r1 Homo sapie	7.94e-03
19	21	2.7	359	48	HUM020E05A	Human fetal brain cDN	7.94e-03
20	21	2.7	386	32	R73112	Yj91a03.sl Homo sapie	7.94e-03
21	21	2.7	388	3	T65442	YC73h07.sl Homo sapie	7.94e-03
22	21	2.7	391	14	T07602	EST05492 Homo sapiens	7.94e-03
23	21	2.7	398	72	H85627	ys88b03.r1 Homo sapie	7.94e-03
24	21	2.7	409	149	W13483	ma84h03.r1 Soares mou	7.94e-03
25	21	2.7	423	9	T87336	yd89g04.sl Homo sapie	7.94e-03
26	21	2.7	432	88	H67117	Yr75a03.sl Homo sapie	7.94e-03
27	21	2.7	433	103	N70544	za83c10.sl Homo sapie	7.94e-03
28	21	2.7	438	5	T71502	Yd35h06.r1 Homo sapie	7.94e-03
29	21	2.7	440	191	AA157291	z052g07.sl Strata gene	7.94e-03
30	21	2.7	440	90	N25934	yx87d01.sl Homo sapie	7.94e-03
31	21	2.7	448	164	C06482	similar to none.	7.94e-03
32	21	2.7	450	113	W01392	Yy67h09.r1 Soares mul	7.94e-03
33	21	2.7	468	61	H15038	Ym20a08.sl Homo sapie	7.94e-03
34	21	2.7	504	27	R55690	Yg88f10.r1 Homo sapie	7.94e-03
35	21	2.7	593	114	W06912	za91g10.r1 Soares fet	7.94e-03
36	20	2.6	236	123	HSAB2C011	H. sapiens partial cD	1.01e-01
37	20	2.6	261	24	R46924	Yg32f07.sl Homo sapie	1.01e-01
38	20	2.6	266	78	R96312	Yq36c09.sl Homo sapie	1.01e-01
39	20	2.6	361	88	H65835	Yr71a03.sl Homo sapie	1.01e-01
40	20	2.6	389	96	N59598	Yy57e09.sl Homo sapie	1.01e-01
41	20	2.6	400	73	H87738	Yy57f01.sl Homo sapie	1.01e-01
42	20	2.6	404	56	T24358	crs1485 Ricinus commu	1.01e-01
43	20	2.6	420	144	N99094	zb81a02.sl Soares sen	1.01e-01
44	20	2.6	491	53	R1C510918A	Rice cDNA, partial se	1.01e-01
45	20	2.6	532	134	N96507	21198 Arabidopsis tha	1.01e-01

ALIGNMENTS

RESULT 1 R95589 337 bp mRNA EST 06-SEP-1995
LOCUS SMNHADA016014T3 Schistosoma mansoni cDNA 5'
DEFINITION R95589
ACCESSION g975959
NID EST.
KEYWORDS blood fluke primer=T3 library-SmAW strain=Egyptian

vector-Bluescript host-Golden Hamster Rsite1-EcoRI Rsite2-XhoI total mRNA was isolated from adult worms using the acid guanidine thiocyanate method (Chomczynski and Sacchi, 1987). Poly (A) + RNA was purified by affinity chromatography on oligo dT column (Aviv and Leder, 1972). The cDNA expression library was constructed using the Lambda zap cloning system (Stratagene-[Uni-Zap XP GigaPack II cloning kit]. The library was amplified (end titre=0.5 x 10E8 pfu/ml) and its efficiency tested for using known S.mansoni probes (chlorion and tropomyosin genes). The size range of inserts for the library was 0.5Kb - 2.5 Kb, the cDNA being directionally cloned between the EcoRI and the XhoI sites of the Bluescript phagemid. Bluescript phagemid was excised by mass in vivo excision of the library and inserted into E.coli XL1 Blue. The plasmid was purified using the Qiagen QiaWell 8 system and sequenced with Taq dye-terminator chemistry (SK73 and T7 primers) for ABI 373A. The library was prepared by Hanan Abdel Hamed, Ain Shams Univ., Cairo and provided by Mohamed Saber, Theodore Bilharz Research Institute, Cairo.

Schistosoma mansoni
Eukaryotes; mitochondrial eukaryotes; Metazoa; Platyhelminthes;
Trematoda; Digenea; Strigeida; Schistosomatoidea;
Schistosomatidae; Schistosoma.
1. (bases 1 to 337)
Ridgers, I.L.
Unpublished (1995)

Contact: Ridders, I.L.
Experimental Taxonomy Unit, Zoology
The Natural History Museum
Cromwell Road, South Kensington, London, SW7 5BD, U.K.
Tel: 0171 0092027

Fax: 01719388754
Email: I.Ridgers@nhs.uk
Schistosoma mansoni cDNA clone SMNHADA016014T3.

```
Location/Qualifiers
1..337
/organism="Schistosoma mansoni"
/strain="Egyptian"
<1..>337
```

```

85 a      58 c      56 g      121 t      17 others
/strain="Egyptian"
<1..>337
3.08; Score 23; DB 76; Length 337;
Similarity 76.58;
26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

tttctnaaaacttcctaagtngngnttcttgta 241
 ||| ||||| || ||| |||||
 TTCTGAAACTTCACGTGAAGGCTTCTTGTA 348

Accession	Length	Source	EST	Accession
AA039422	439 bp	pregnant uterus	NbhpU	30-AUG-1996
K39d04.r1		Soares	Homo	
85191.5				
AA039422				
1515699				
St.				
uman.				

omo sapiens
karyotae; mitochondrial eukaryotes; Metazoa; Chordata;
ertebrata; Eutheria; Primates; Catarrhini; Homi-
(bases 1 to 439)

Miller, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Goldman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Marston, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevisan, E., Waterston, R., Williamson, A., Woldmann, P. and
 Ziegler, W.

the WashU-Merck EST Project
unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lln.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 246.

FEATURES
source

```

source
I...439
/organism="Homo sapiens"
/organ="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
Site_2: Eco RI; Site_3: 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AACTCGAGAAATCGGCGCGCTTTTTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
/clone="485191"
/clone_lib="Soares pregnant uterus NbHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
<1..>439
125 a 81 C 79 g 145 t 9 others

mrna
BASE COUNT
ORIGIN
Query Match 3.0%; Score 23; DB 159; Length 439;
Best Local Similarity 71.7%; Pred. No. 3.49e-05;
Matches 38; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 234 ttgtgctgctaaagtatttacttctattagctgttctgacctta 286
||||| ||||||| ||||||| ||||||| ||| |||||
OY 64 TTGTGTGCTGCTAAAGTTTATAGTCAATAGGTGCTAAATAGTATTACTTA 116
||||| ||||||| ||||||| ||||||| ||| |||||

```

RESULT	3
LOCUS	N86668
DEFINITION	J9617F Fetal heart, Lambda ZAP Express Homo sapiens CDNA clone J9617 5', similar to CAYEOLIN.
ACCESSION	N86668
NID	G1439870
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	Liew,C.C.
TITLE	cDNAs from fetal heart (1996)
JOURNAL	Unpublished (1996)

Contact: Liew CC
Molecular Cardiology
University of Toronto
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
Tel: 4169789758
Fax: 4169785650
Email: liewccc@utcc.utoronto.ca
Seq primer: GAAATTAACCGCTCACTAAAGGG.

FEATURES	SOURCE
-----------------	---------------

```

/organism="Homo sapiens"
/vector: Lambda ZAP Express: Site_1: EcoRI; Site_2:
XhoI. mRNA was purified from human fetal hearts (8-10
weeks). mRNA was synthesized using a XhoI-Oligo dT
adaptor-primer. EcoRI adaptors were ligated, followed by
digestion with XhoI. For directional cloning into
predesigned lambda ZAP Express.
/clone="19617"

```


COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 353.
Location/Qualifiers

FEATURES

source
1..604
/organism="Homo sapiens"
/note="Vector: p7T3D (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAGTGGGAGCGCGCGAGTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p7T3 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo. RNA from normal foreskin melanocytes
(FS374) was kindly provided by Dr. Anthony P. Albino.
/clone="291949"
/clone_lib="Soares melanocyte 2NBHM"
/sex="Male"
/lab_host="DH10B (ampicillin resistant)"
<1..>604

BASE COUNT
ORIGIN

166 a 149 c 126 g 152 t 11 others

Query Match

Best Local Similarity 2.9%; Score 22; DB 113; Length 604;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 458 caagctcttactggagcattctgaacttaggaataac 497

Cp 145 CAAGCTCTTACGGCTACGTTCTTACGGTAACTAATAC 106

RESULT 9

LOCUS H85028 132 bp mRNA EST 14-NOV-1995
DEFINITION y888q01.s1 Homo sapiens cDNA clone 221904 3'.
ACCESSION H85028
NID g1064730

KEYWORDS

EST.
human clone-221904 primer-Promega -21ml3 library-Soares retina
N2b5HR vector-p7T3D (Pharmacia) with a modified polylinker
host-DH10B (ampicillin resistant) Rsite1-Not I Rsite2-Eco RI 1st
strand cDNA was primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAGTGGGAGCGCGCGAGTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI adapters
(Pharmacia), digested with Not I and cloned into the Not I and Eco
RI sites of a modified p7T3 vector (Pharmacia). The retinas were
obtained from a 55 year old Caucasian male and total cellular
poly(A)⁺ RNA was extracted 6 hrs after their removal. The retina
RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento Soares and
M.Fatima Bonaldo.

ORGANISM

Homo sapiens
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Homindae; Homo.
1 (bases 1 to 132)

REFERENCE

AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
The WashU-Merck EST Project

TITLE

JOURNAL
COMMENT

Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 73
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES

source
1..132
/organism="Homo sapiens"
/clone="221904"
<1..>132
mRNA
BASE COUNT 51 a 15 c 23 g 39 t 4 others
ORIGIN

Query Match

Best Local Similarity 2.7%; Score 21; DB 72; Length 132;
Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Db 52 atgttaatactgataaaacacatgttcaatcatcgnaa 91

QY 1 ATGTTAAATCTTGAAACAAACATATATGTCATCATGGGAA 40

RESULT 10

LOCUS T47205 172 bp mRNA EST 08-FEB-1995
DEFINITION yb53g07.r1 Homo sapiens cDNA clone 74940 5'.

ACCESSION T47205

NID 9649187

KEYWORDS

EST.
SOURCE human clone-74940 library-Stratagene ovary (#937217)
vector=Bluescript SK host-SOLR cells (kanamycin resistant)
Primer-M13RP1 Rsite1-EcoRI Rsite2-XhoI Cloned unidirectionally.
Primer: Oligo dT. Total ovary tissue, normal 49 year old caucasian
female. Average insert size: 0.8 kb; Uni-ZAP XR Vector; 5' adaptor
sequence: 5'-GATTTCGGCAGCAG-3'; 3' adaptor sequence:
5'-CTCGAGTTTCTTTTCTTTT-3'.

ORGANISM

Homo sapiens
Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;
Eutheria; Primates; Catarrhini; Homindae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 172)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E.,
Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.

TITLE

WashU-Merck EST Project

JOURNAL

Unpublished (1995)

COMMENT

Other ESTs: yb53g07.s1
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Putative full length read.

FEATURES

source
1..172
/organism="Homo sapiens"
/clone="74940"
BASE COUNT 48 a 30 c 30 g 64 t
ORIGIN

Query Match

2.7%; Score 21; DB 17; Length 172;

Best Local Similarity 70.6%; Pred. No. 7.94e-03;
Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

	Matches	36	Conservative	0	Mismatches	15	Indels	0	Gaps	0
Db	98	tttggatagttgctttcttgattgattcttttctctogaacttttaa	148							
Cp	193	tttgataaagtcgcttctgggtgaattaatgttctaatatattttcaa	143							

RESULT	11			
LOCUS	AA092855	210 bp	mRNA	EST
DEFINITION	l19883 seq. F Fetal heart, Lambda ZAP Express Homo sapiens CDNA 5'.			
ACCESSION	AA092855			
NID	g1637868			
KEYWORDS	EST.			

SOURCE	DOI: human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 210)
AUTHORS	Liew,C.C.
TITLE	CDNAs from fetal heart (1996)
JOURNAL	Unpublished (1996)

Contact: Liew CC
 Molecular Cardiology
 University of Toronto
 Banking Institute, 100 College St., Toronto, Ontario, M5G1L5
 Tel: 4169788758
 Fax: 4169785650
 Email: liewcc@utcc.utoronto.ca
 PCR Primers
 FORWARD: 5' GCCAGCTCGAAATTAAACCTCACTAAAGG 3'
 BACKWARD: 5' CCAGTGAATTGTAATACGACTCACTATAGGCG 3'
 Seq primer: 5' GAAATTAACCTCACTAAAGG 3'.
 Location/Qualifiers

```

source
1..210
/organism="Homo sapiens"
/note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2:
XhoI; mRNA was purified from human fetal hearts (0-10
weeks). cDNA was synthesized using a XhoI-Oligo dT
adaptor-primer. EcoRI adaptors were ligated, followed by
digestion with XhoI, for directional cloning into
predigested lambda ZAP Express."
/clone_lib="Fetal heart, Lambda ZAP Express"
/lab_host="E. coli XL1-Blue"
<1..>210
78 a 45 c 39 g 48 t
mRNA
BASE COUNT

```

```

every Match          2.78;      Score 21;  DB 182;  Length 210;
Best Local Similarity 86.23;    Pred. No. 7.94e-03;
Matches 25;  Conservative 0;  Mismatches 4;  Indels 0;  Gaps 0;

```

Db	84	tctgaaccacaattttctcaaaaccagtaat	112
	248	TCCTTACCAATTTCGTCAAAACCATTAAT	220
Cp			

RESULT	12				
LOCUS		HUM003B01A	214 bp	mRNA	EST
DEFINITION		Human fetal brain CDNA 3'-end			26-AUG-1995
ACCESSION		D59261			
INSD		9960367			
KEYWORDS		EST(expressed sequence tag); Human fetal brain; similar to none(May 29, 1995).			
SOURCE		<p> Homo sapiens (library: Clontech human fetal brain polyA+ mRNA (#6535)) cDNA to mRNA. </p>			

ORGANISM	REFERENCE AUTHORS
Homo sapiens	Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M., 1 (bases 1 to 214)
Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	

Takaichi, A., Takeda, S., Watanabe, T., Takahashi, E.-I., Hirai, Y.,
Maekawa, H., Shin, S. and Nakamura, Y.
Unpublished (101)
Unpublished (1995)
Submitted (30-May-1995), to DDBJ by:
Tsutomu Fujiwara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co., Ltd
463-10 Kagasuno Kawauchi-cho
Tokushima, Tokushima
771-01

Phone: 0886-65-2888
Fax : 0886-37-1035.

max . 0000 37-1033.	Location/Qualifiers
FEATURES	1..214
source	

BASE COUNT	35 a	32 c	54 g	54 t	39 others
ORIGIN					
	/organism="Homo sapiens"				
	/clone_lib="Clontech human fetal brain polyA+ mRNA				
	(#6595)"				

Query Match	2.7%	Score 21;	DB 48;	Length 214;
Best Local Similarity	42.6%;	Pred. No. 7.94e-03;		
Matches	23; Conservative	14; Mismatches	17; Indels	0; Gaps
				0;

```
6 dgsrgggggggtttvtbyattgadddaddtttadaarrgktgnwtgttha 59
   : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
598 GGTGTTGGTGTTTCAATACAACTTCCTTAAGAAATCGAAGAGCGTGCACCTTTA 651
```

RESULT	13	LOCUS	HUM238E03B	233 bp	mrna	EST	29-AUG-1995
DEFINITION			Human aorta cDNA 5'-end	GEN-238E03.			
ACCESSION			D62108				
NID			q965884				

KEYWORDS EST(expressed sequence tag); Human aorta; similar to none(May 29,1995).

SOURCE Homo sapiens (library: Clontech human aorta polyA+ mRNA (#6572)) cDNA to mRNA.

ORGANISM
GENE TO NANA.
HOMO sapiens
Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarhini; Hominoidea; Homo.
1 (bases 1 to 233)
Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagai,
Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H.,
Takaichi, A., Takeda, S., Watanabe, T., Takahashi, E.-I., Hirai,
Maezawa, H., Shin, S. and Nakamura, Y.

TITLE	COMMENT
Maekawa, H., Shin, S. and Nakamura, Y.	
Unpublished (303)	
Unpublished (1995)	
Submitted (30-May-1995)	to DDBJ by:

Tsutomu Fujiwara
Osaka GEN Research Institute
Osaka Pharmaceutical Co., Ltd
463-10 kagasuno Kawauchi-cho
Tokushima, Tokushima
771-01

Japan
Phone: 0886-65-2888
Fax : 0886-37-1035

FEATURES
source

ASE COUNT	91 a	34 c	32 g	69 t	7 others
1.1255	/organism="Homo sapiens"				
	/clone_lib="Clontech human aorta polyA+ mRNA (#572)"				

Query Match 2.7%; Score 21; DB 48; Length 233;
Best Local Similarity 71.8%; Pred. No. 7.94e-03;
Matched 35

[illegible]

human clone-116989 library-Scars fetal liver spleen INFLS vector-p773D (Pharmacia) with a modified polylinker host=PH10B (ampicillin resistant) primer-M13AP1 Rsite1-Pac I Rsite2-Eco RI Liver and spleen from a 20 week-post conception male fetus-1st strand cDNA was primed with a Pac I - oligo(dT) primer [5]

Search completed: Fri Dec 5 08:08:05 1997
Job time : 502 secs.

WIRE

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
on: Fri Dec 5 08:08:27 1997; MasPar time 252.09 Seconds
Similar output not generated. 840.498 Million cell updates/sec

Title: >US-08-790-043A-2
Description: (1-771) from US08790043A.seq
Perfect Score: 771
N.A. Sequence: 1 ATGTTAAATCTTGAACAA.....GATTCACGCAATTAATAAA 771
Comp: TACAATTAGACTTTGTT.....CTAAGTGGTTAAATTATT

Scoring table: TABLE default
Gap 6
Nmatch STD : Dbase 0; Query 0
Searched: 359085 seqs, 137405154 bases x 2
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: EST-STS-THREE
1:EST199 2:EST200 3:EST201 4:EST202 5:EST203 6:EST204
7:EST205 8:EST206 9:EST207 10:EST208 11:EST209 12:EST210
13:EST211 14:EST212 15:EST213 16:EST214 17:EST215
18:EST216 19:EST217 20:EST218 21:EST219 22:EST220
23:EST221 24:EST222 25:EST223 26:EST224 27:EST225
28:EST226 29:EST227 30:EST228 31:EST229 32:EST230
33:EST231 34:EST232 35:EST233 36:STS1 37:STS2 38:STS3
39:STS4 40:STS5 41:STS6 42:STS7 43:STS8 44:STS9 45:STS10
46:STS11 47:STS12 48:STS13
EST-STS-FOUR
49:gnEST1 50:gnEST2 51:gnEST3 52:gnEST4 53:gnEST5
54:gnEST6 55:gnEST7 56:gnEST8 57:gnEST9 58:gnEST10
59:gnEST11 60:gnEST12 61:gnEST13 62:gnEST14 63:gnEST15
64:gnEST16 65:gnEST17 66:gnEST18 67:gnEST19 68:gnEST20
69:gnEST21 70:gnEST22 71:gnEST23 72:gnEST24 73:gnEST25
74:gnSTS 75:gnEST1 76:gnEST2 77:gnEST3 78:gnEST4
79:gnEST5 80:gnEST6 81:gnEST7 82:gnEST8 83:gnEST9
84:gnEST10 85:gnEST11 86:gnEST12 87:gnEST13 88:gnEST14
89:gnEST15 90:gnEST16 91:gnEST17 92:gnEST18 93:gnEST19
94:gnEST20 95:gnEST21 96:gnEST22 97:gnEST23 98:gnEST24
99:gnEST25 100:gnEST26 101:gnEST27 102:gnEST28
103:gnEST29 104:gnEST30 105:gnEST31 106:gnSTS 107:ueEST1
108:ueEST2

Statistics: Mean 10.599; Variance 2.102; scale 5.042
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	23	3.0	396	25	N59624	yy74c01.s1 Soares fet	2.18e-05
2	22	2.9	233	85	HS1166343	zsl0b06.r1 Soares NbH	3.39e-04
3	22	2.9	233	71	AA251633	zsl0b06.r1 Soares NbH	3.39e-04
4	22	2.9	393	49	AA125483	mq71c01.r1 Stratagene	3.39e-04
5	22	2.9	393	7	AA125483	mq71c01.r1 Stratagene	3.39e-04
6	22	2.9	424	50	AA138421	mq88b09.r1 Stratagene	3.39e-04
7	22	2.9	427	71	AA251305	zsl0b06.s1 Soares NbH	3.39e-04
8	22	2.9	427	85	HS1166232	zsl0b06.s1 Soares NbH	3.39e-04
9	21	2.7	333	32	AA072168	mm69c03.r1 Stratagene	4.77e-03
10	21	2.7	341	108	MM1139719	my25c01.r1 Barstead m	4.77e-03
11	21	2.7	341	52	AA221186	1470C3 czapFFD2.1, D	4.77e-03
12	21	2.7	358	1	N97897	MAAD0115.M2R Schistos	4.77e-03
13	21	2.7	394	13	AA115832	zf63a01.r1 Soares ret	4.77e-03
14	21	2.7	408	30	AA059032	ma84h03.r1 Soares mou	4.77e-03
15	21	2.7	409	95	MM4838	144 Mouse VM cDNA lib	4.77e-03
16	21	2.7	415	27	W36178	mul5a09.r1 Soares 2Nb	4.77e-03
17	21	2.7	449	18	AA200846	mul5a09.r1 Soares 2Nb	4.77e-03
18	21	2.7	449	61	AA200846	mul5a09.r1 Soares 2Nb	4.77e-03
19	21	2.7	449	104	MMAA846	zq53b05.s1 Stratagene	4.77e-03
20	21	2.7	593	19	AA205935	mo1a05.r1 Stratagene	5.99e-02
21	20	2.6	135	55	AA088037	D. melanogaster STS d	5.99e-02
22	20	2.6	193	36	DM86D10S	human STS A001140.	5.99e-02
23	20	2.6	196	42	G19638	human STS A001140.	5.99e-02
24	20	2.6	250	85	HS1166217	zsl0a01.s1 Soares NbH	5.99e-02
25	20	2.6	250	71	AA251371	human STS SHGC-5955 c	5.99e-02
26	20	2.6	328	39	G10965	human STS WI-30760.	5.99e-02
27	20	2.6	348	42	G22403	SWAMCAL059SK Bruglia m	5.99e-02
28	20	2.6	348	1	W84913	human STS SHGC-12056	5.99e-02
29	20	2.6	400	41	G17808	ze59g06.s1 Soares ret	5.99e-02
30	20	2.6	412	21	AA019095	mv52f04.r1 Soares mou	5.99e-02
31	20	2.6	422	64	AA230876	mv52f04.r1 Soares mou	5.99e-02
32	20	2.6	422	99	MMAA30876	human STS WI-14071.	5.99e-02
33	20	2.6	435	42	G21132	ms21g02.r1 Stratagene	5.99e-02
34	20	2.6	445	51	AA169090	ze03h12.s1 Soares fet	5.99e-02
35	20	2.6	463	2	W94669	ze03h12.s1 Soares fet	5.99e-02
36	20	2.6	463	90	HSW6694	z16g10.s1 Soares fet	5.99e-02
37	20	2.6	470	30	AA055104	5770 Arabidopsis thal	5.99e-02
38	20	2.6	473	77	AT5074	z55e09.r1 Stratagene	5.99e-02
39	20	2.6	486	22	AA211548	21198 Arabidopsis tha	5.99e-02
40	20	2.6	532	77	AT50720	va59g01.r1 Soares mou	5.99e-02
41	20	2.6	559	101	MMAA39917	z131g06.s1 Soares pre	5.99e-02
42	20	2.6	585	9	AA131271	CK01713.Sprime Drosop	5.99e-02
43	20	2.6	676	34	AA141420	human STS WI-7527.	5.99e-02
44	20	2.6	1990	38	G06577	zr95d02.s1 Soares NbH	6.65e-01
45	19	2.5	288	54	AA215434		

ALIGNMENTS

RESULT	1	N59624	396 bp	mrna	EST	28-JAN-1997
LOCUS		yy74c01.s1	Soares fetal liver spleen	INFLS	Homo sapiens	cDNA clone
DEFINITION		248448 3'				
ACCESSION		N59624				
NID		g1203514				
KEYWORDS		human.				
SOURCE		Homo sapiens				
ORGANISM		Eukaryotae; Mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE		1 (bases 1 to 396)				
AUTHORS		Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.				
TITLE		WashU-Merck EST Project				
JOURNAL		Unpublished (1995)				
COMMENT		Contact: Willson RK WashU-Merck EST Project Washington University School of Medicine				

ORIGIN

Query Match 2.9%; Score 22; DB 71; Length 233;
 Best Local Similarity 80.0%; Pred. No. 3.39e-04;
 Matches 36; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

Db 69 gattgcttaatttattgtaacaca-atacatttgcact 112
 ||| ||||| ||||| ||||| ||| ||||| |||||

Cp 633 GATTCTTTGAAGTGTATTGAAACACACCCACCTTTTGCAC 589

RESULT

4 AAL125483 393 bp mRNA EST 09-FEB-1997
 LOCUS mg71c01.r1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
 DEFINITION clone 584160 5' similar to SW:DLDH_BACSU P21880 LIPOAMIDE
 DEHYDROGENASE COMPONENT ;

ACCESSION

AAL125483

NID

g1683736

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

1 (bases 1 to 393)
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,F., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 The WashU-HHMI Mouse EST Project
 Unpublished (1996)

TITLE

JOURNAL

COMMENT

Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:358808
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 382.
 Location/Qualifiers

FEATURES

source

1..393
 /organism="Mus musculus"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; Cloned unidirectionally. Primer: Oligo dt. From M2
 cells, a highly metastatic derivative of the K-1735
 (mouse) melanoma. Average insert size: 1.0 kb; Uni-ZAP XR
 Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3'
 adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'
 /clone="584160"
 /clone_lib="Stratagene mouse melanoma (#937312)"
 /dev_stage="M2 cells"
 /lab_host="SOLR (kanamycin resistant)"
 <1..>393

BASE COUNT

ORIGIN

Query Match 2.9%; Score 22; DB 49; Length 393;
 Best Local Similarity 84.4%; Pred. No. 3.39e-04;
 Matches 27; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 312 ttgttttggtagaactacaatttagatcaa 343
 ||| ||||| ||| ||| ||||| |||||

Qy 59 TTGCTTTTGGTGTGCTAAAGTTTATGATCAA 90

RESULT

LOCUS

5 AAL125483 393 bp mRNA EST 22-NOV-1996

DEFINITION

mg71c01.r1 Stratagene mouse melanoma (#937312) Mus musculus cDNA
 clone 584160 5' similar to SW:DLDH_BACSU P21880 LIPOAMIDE
 DEHYDROGENASE COMPONENT ;

ACCESSION

NID

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

Db 51 tgcagaaagagtgaagctttggagtcagccctgaaagcacaagaataatgcacctcg 110
||||| |||| | |||| | |||| | |||| | |||| |
QY 591 TGCAAAAGGTGTGGGTCGTTTCAATACAAATCTTAAGCAAATCGAAGACCGTGACCTTT 650

NID
KEYWORDS
918401/2
EST.

```

Db    111 agacgcgtaaacgctatcatcagaagaagtga 139
      | | | | | | | | | | | | | | | | |
QY    651 AAAACGTAACGTTGATCAAGTAGAAGTAG 679

RESULT 10
ID     MM1139719 standard; RNA; EST; 341 BP.
AC     AA221186;
NI     G1840172
DT     14-FEB-1997 (Rel. 50, Created)
DT     14-FEB-1997 (Rel. 50, Last updated, Version 1)
DE     my25c01.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
DE     clone 696864 5'.
DE     KW
OS     Mus musculus (house mouse)
OC     Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
OC     Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
OC     Mus.
RN     [1]
RP     1-341
RA     Marra M., Hillier L., Allen M., Bowles M., Dietrich N.,
RA     Dubuque T., Geisel S., Kucaba T., Lacy M., Le M., Martin J.,
RA     Morris M., Schellenberg K., Steptoe M., Tan F., Underwood K.,
RA     Moore B., Theising B., Wylie T., Lennon G., Soares B., Wilson R.,
RA     Waterston R.;
RT     "The WashU-HHMI Mouse EST Project";
RL     Unpublished.
CC     Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project
CC     Washington University School of Medicinep 4444 Forest Park Parkway,
CC     Box 8501, St. Louis, MO 63108 tel: 314 286 1800 fax: 314 286 1810
CC     Email: mouse@et.watson.wustl.edu This clone is available
CC     royalty-free through LLNL ; contact the IMAGE Consortium
CC     (info@image.llnl.gov) for further information. MGI:430424 Seq
CC     primer: -28ml3 rev2 ET from Amersham High quality sequence stop:
CC     234.
FH     Key
FH     Location/Qualifiers
FT     source
FT     1..341
FT     /organism="Mus musculus"
FT     /strain="FVB/N"
FT     /note="Vector: pT7T3D-pac (Pharmacia) with a modified
FT     polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
FT     was primed with a Not I - oligo(dT) primer [5',
FT     TGTTACGAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTTTTTTTTTT
FT     3']; double-stranded cDNA was ligated to Eco RI adaptors
FT     [GTTGATTCGTACC], digested with Not I and cloned into the
FT     Not I and Eco RI sites of the modified pT7T3 vector.
FT     Library constructed by Bob Barstead."
FT     /clone_lib="Barstead mouse pooled organs MPLRB4"
FT     /sex="mixed"
FT     /dev_stage="7 day"
FT     /lab_host="DH10B"
FT     <1..>341
FT     mRNA
FT     Sequence 341 BP; 101 A; 67 C; 73 G; 100 T; 0 other;

Query Match 2.7%; Score 21; DB 108; Length 341;
Best Local Similarity 81.8%; Pred. No. 4.77e-03;
Matches 27; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db    269 acaattgttaataaacattagaagaatgtcttga 301
      | | | | | | | | | | | | | | | | |
Cp    383 ACAATGTTAATGATGAAGAACTAATGCTCTGA 351

RESULT 11
LOCUS   AA221186 341 bp mRNA EST 12-FEB-1997
DEFINITION mv25c01.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA
ACCESSION AA221186
NID       G1840172
KEYWORDS  EST.

```


Db 285 aaaaatgatgaagtgcctgggttgctggatttggctgaaaaaggtcatgctgtaggtgatattc 344
 QY 203 AAAGCCATGAAGAGGTTATTAATGGTTTTTGAGCAAAATGGTAAAGATGTTGGCAATATTG 262
 Db 345 ctggagt 351
 QY 263 ATGGTGT 269

Search completed: Fri Dec 5 08:12:53 1997
 Job time : 266 secs.